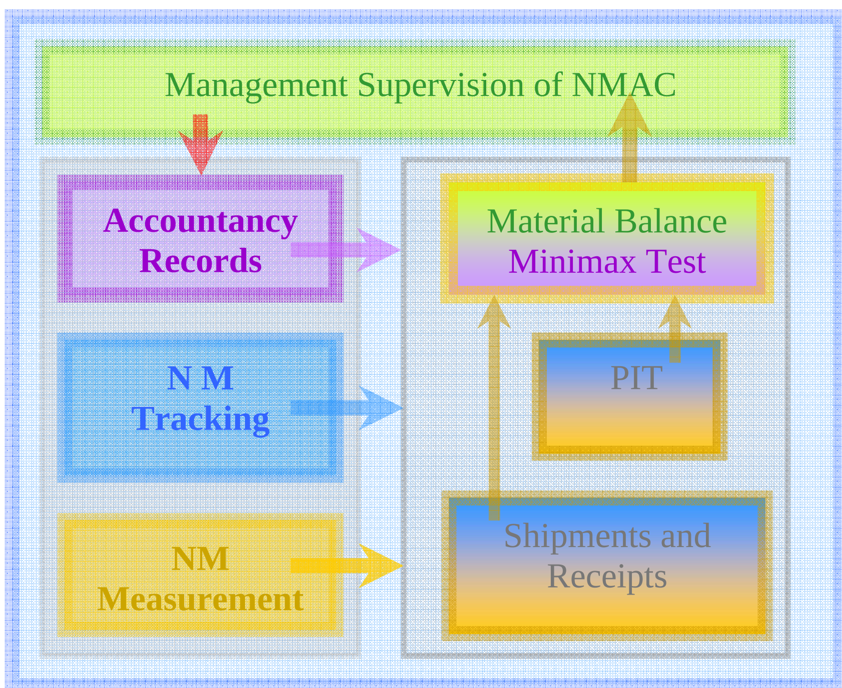
JRC Scientific and Technical Reports

MINIMAX TESTS OF COMPOSITE NULL HYPOTHESES APPLIED TO NUCLEAR MATERIAL BALANCES

M. T. Franklin



EUR 23800 EN - March 2009

The Institute for the Protection and Security of the Citizen provides research-based, systems-oriented support to EU policies so as to protect the citizen against economic and technological risk. The Institute maintains and develops its expertise and networks in information, communication, space and engineering technologies in support of its mission. The strong cross-fertilisation between its nuclear and non-nuclear activities strengthens the expertise it can bring to the benefit of customers in both domains.

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen

Contact information
Address: Nuclear Security Unit, IPSC, CCR, 21025 Ispra(VA), Italy
E-mail: JRC-NUSEC-SECRETARIAT@ec.europa.eu
Tel.: +390332789603
Fax: +390332789185

http://ipsc.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC51379

EUR 23800 EN

ISSN 1018-5593

Luxembourg: Office for Official Publications of the European Communities

Printed in Italy

MINIMAX TESTS OF COMPOSITE NULL HYPOTHESES APPLIED TO NUCLEAR MATERIAL BALANCES

Contents

Contents (continued)

MINIMAX TESTS OF COMPOSITE NULL HYPOTHESES APPLIED TO NUCLEAR MATERIAL BALANCES

1. Introduction and Summary of Report

This first section provides an introduction to the problem of assessing material balances in bulk handling facilities. It then sets up the minimax formulation of balance assessment as a decision problem with loss function. It then provides a preview of the findings and a guide to the contents of sections 2 to 5 of the report.

1.1 Introduction

In a facility processing nuclear material, assessing the material balance at regular intervals is an essential part of assuring that all material is accounted for. The material balance (MUF)[1] can be treated as a Gaussian random variable since the balance includes the cumulative effect of the many legitimate measurement errors that are in the accounts. The measurement error standard deviation of MUF is denoted σ_{MUF} . The value of σ_{MUF} is one of the factors determining the range of MUF values that will be considered as acceptable. In computing σ_{MUF} , the statistical approach makes the assumption that the accountancy values are all based on measured values coming from correctly executed measurement procedures i.e. without **without human errors**. Starting from this assumption and using information about how each measurement has been made, the statistical approach computes a standard deviation of MUF. This standard deviation reflects the probability distribution which the MUF would have due to those components of measurement error that have probability distributions. Computing σ_{MUF} takes no account of possible error components that do not have a probability distribution (biases[2]). Provided there are no biases in measurement procedures and provided the accounting information includes all the material relevant to the balance, the operator's MUF is "an

[1] The material balance value is often referred to as "material unaccounted for" and denoted MUF.

[2] Error components that do not have a probability distribution are referred to as biases.

observation" on a probability distribution with zero expected value and with standard deviation given by σ_{MUF} .

In this report we consider the balance problem when the situation is not as simple as that described above. Here we consider the problem of assessing the acceptability of a material balance when some small amount of well identified material has not been measured. This kind of situation can occur when all the material in a process area cannot be measured. Typical examples occur in the case of reprocessing plants and enrichment facilities. The heel of a tank in a reprocessing facility may contain a small amount that is difficult to measure. The cascade area in an enrichment facility may contain the usual process hold-up amounts if the balance is being assessed while the process remains in operation.

In such situations, it is useful that the balance be assessed for acceptability while knowing that there are some small amounts of hold-up which have not been measured. Unmeasured hold-up is not included in the balance computation and hence the balance is an **incomplete balance**. This incomplete balance is the sum of a mean value (not necessarily zero because the hold-up is not accounted for) plus an accumulation of measurement error intrinsic to the accounting values of the material that has been measured.

In many cases it is possible to establish upper and lower acceptability limits for the true amount of material in hold-up. Such limits are not an estimate of what is contained in the hold-up instead they are a statement of what values are to be considered acceptable. These limits can be derived taking account of the processing history generating the hold-up. The limits then determine what is an acceptable range for the mean value of MUF and this range becomes a **composite null hypothesis** for testing the incomplete material balance. When the composite null hypothesis is true, the incomplete balance is the sum of a mean value contained between upper and lower bounds plus an accumulation of measurement error intrinsic to the accounting values of the material that has been measured. As before, if we know the probability distribution of the measurement errors, the standard deviation of the measurement error in MUF can be computed and will be denoted σ_{MUF} . This value is determined from the measurement history of the material that has been processed during the balance period along with the probability distributions of the measurement errors.

The concept of composite null hypothesis has another application in the assessment of nuclear material balances. Even when all material in the balance accounts has a measured value that was established using correct procedures, there may be uncorrected biases. The measurement specialists can be aware that such biases can exist and that their existence is difficult to establish. They may consider that the possibility of a small bias should be allowed for in assessing the material balance. This can be achieved by using a composite null hypothesis when it is possible to establish a range of values that can be tolerated for the cumulative effect of bias in the material balance.

In this report we describe simple minimax tests for assessing the acceptability or not of a material balance when there is a composite null hypothesis. The statistical procedure compares the balance value to test acceptance limits that are determined from,

1) the standard deviation (σ_{MUF}) of the cumulative contribution of measurement errors in the balance. In this report we consider that σ_{MUF} is a known value.

2) the management or inspectors aversion to false alarms. This aversion is represented as a requirement to choose the statistical test rule so that the maximum false alarm probability is equal to a desired target value.

3) Upper and lower tolerance limits for any hold-up amounts that have not been included in the material balance.

4) Upper and lower limits for the effect (on MUF) of any uncorrected biases in the measurement processes.

1.2 The Minimax Formulation

From now on we use the following notation:

- X denotes the incomplete material balance (MUF),
- θ denote the mean value of X where $-\infty < \theta < +\infty$,
- σ or σ_{MUF} denotes the known standard deviation of X,

- $X \sim N(\theta,\sigma)$ denotes " X has a Gaussian distribution with mean θ and standard deviation σ" .
- The closed interval $\left[a\sigma, b\sigma\right]$ where $a \le b$, denotes the range of acceptable values of θ. This range of values will also be denoted $\mathbf{H_0}$.
- The null hypothesis is denoted $\theta \in \left[a\sigma, b\sigma\right]$ or $\theta \in H_0$. The null hypothesis is called composite when $a < b$ and is called simple when $a = b$.
- The alternative hypothesis is $\theta \notin \left[a\sigma, b\sigma\right]$ or $\theta \notin H_0$
- Accept H_0 (or $\mathbf{AH_0}$) denotes the action "Accept the hypothesis $\theta \in H_0$"
- Reject H_0 (or $\mathbf{RH_0}$) denotes the action "Reject the hypothesis $\theta \in H_0$"
- $L(\theta, \mathbf{AH_0})$ and $L(\theta, \mathbf{RH_0})$ denote the loss when the value of the mean is θ and the action taken is Accept H_0 or Reject H_0 respectively.
- $\mathbf{d}$: $\mathrm{IR} \to \{\mathbf{AH_0}, \mathbf{RH_0}\}$ is the decision rule of acceptance or rejection of $\mathbf{H_0}$ based on the value of the material balance, i.e. $x \mapsto \mathrm{d}(x)$.
- $\mathrm{p}_1 > 0$ is the penalty associated with Type 1 error (i.e. cost of false alarm)
- $\mathrm{p}_2 > 0$ is the penalty associated with Type 2 error (i.e. cost of non detection).

The Expected Loss: When θ is true and the decision rule $\mathbf{d}$ is used, the realised loss will be $L(\theta,\mathbf{d}(x))$. The expected loss for any θ is by definition,

$$E_{x|\theta}\left\{L\left(\theta, d(x)\right)\right\} =$$

$$L\left(\theta, AH_0\right)P\left(d(x) = AH_0 \mid \theta\right) + L\left(\theta, RH_0\right)P\left(d(x) = RH_0 \mid \theta\right)$$

Letting $\alpha_d(\theta)$ denote $P\left(d(x)=RH_0|\theta\right)$ and hence having $P\left(d(x)=AH_0|\theta\right) = 1-\alpha_d(\theta)$, we can write the expected loss as

$$E_{x|\theta}\left\{L\left(\theta,d(x)\right)\right\} = L\left(\theta,AH_0\right)\left[1-\alpha_d(\theta)\right] + L\left(\theta,RH_0\right)\alpha_d(\theta)$$

Note that the function $E_{x|\theta}\left\{L\left(\theta,d(x)\right)\right\}$ is sometimes denoted $R\left(\theta,d\right)$ and referred to as the risk function. It gives the "risk equation" for such decisions as :

$$R\left(\theta,d\right) = L\left(\theta,AH_0\right)\left[1-\alpha_d(\theta)\right] + L\left(\theta,RH_0\right)\alpha_d(\theta)$$

The Idea of Minimax

A minimax rule is defined as being a rule **d** that minimises the worst expected loss i.e. worst with respect to θ $[\text{where } -\infty<\theta<+\infty]$. In other words, a minimax rule is one that minimizes,

$$\sup_{-\infty<\theta<+\infty} R\left(\theta,d\right)$$

Note that for any decision rule of acceptance or rejection of a null hypothesis, the supremum of expected loss can be rewritten to give,

$$\sup_{-\infty<\theta<+\infty} R\left(\theta,d\right) = \sup\left\{ \sup_{\theta\in H_0} R\left(\theta,d\right) \,,\; \sup_{\theta\notin H_0} R\left(\theta,d\right) \right\}$$

We now need to specify the loss function that will define $R\left(\theta,d\right)$ in this report i.e. we need to specify L(θ, $\mathbf{AH_0}$) and L(θ, $\mathbf{RH_0}$) .

The Loss Function

The loss function to be used here for developing minimax rules is given by,

$$L(\theta, \mathbf{AH_0}) = \begin{cases} p_2 & \text{for } \theta < a\sigma \\ 0.0 & \text{for } \theta \in [a\sigma, b\sigma] \\ p_2 & \text{for } \theta > b\sigma \end{cases}$$

$$L(\theta, \mathbf{RH_0}) = \begin{cases} 0.0 & \text{for } \theta < a\sigma \\ p_1 & \text{for } \theta \in [a\sigma, b\sigma] \\ 0.0 & \text{for } \theta > b\sigma \end{cases}$$

If this definition of L(θ, a) is substituted into the risk equation we get

$$R(\theta,d) = \begin{cases} p_2\left[1-\alpha_d(\theta)\right] & \text{for } \theta < a\sigma \\ p_1\alpha_d(\theta) & \text{for } \theta \in [a\sigma, b\sigma] \\ p_2\left[1-\alpha_d(\theta)\right] & \text{for } \theta > b\sigma \end{cases}$$

Substituting this expression for $R(\theta,d)$ gives,

$$\sup_{-\infty<\theta<+\infty} R(\theta,d) = \sup\left\{ \sup_{\theta\in H_0} p_1\alpha_d(\theta),\ \sup_{\theta\notin H_0} p_2\left[1-\alpha_d(\theta)\right]\right\}$$

$$\sup_{-\infty<\theta<+\infty} R(\theta,d) = \sup\left\{ p_1 \sup_{\theta\in H_0} \alpha_d(\theta),\ p_2 \sup_{\theta\notin H_0} \left[1-\alpha_d(\theta)\right]\right\}$$

$$\sup_{-\infty<\theta<+\infty} R(\theta,d) = p_1 \sup\left\{ \sup_{\theta\in H_0} \alpha_d(\theta),\ \frac{p_2}{p_1} \sup_{\theta\notin H_0} \left[1-\alpha_d(\theta)\right]\right\}$$

For notation we put $Q = \frac{p_2}{p_1}$; $\alpha_d^* = \sup_{\theta \in H_0} \alpha_d(\theta)$ [3]

and $\beta_d^* = \sup_{\theta \notin H_0} \left[1 - \alpha_d(\theta)\right]$. [4]

Using this notation we have,

$$\sup_{-\infty < \theta < +\infty} R(\theta, d) = p_1 \sup\left\{ \alpha_d^* , Q\beta_d^* \right\}$$

This expresses the supremum of the expected loss associated with **d** in terms of the worst false alarm probability, the worst detection probability and the penalty costs of the two error types. We now wish to find the decision rule **d**: IR(X) → {$\mathbf{AH_0}$, $\mathbf{RH_0}$} that will minimise this expression. Such a rule will be referred to as the minimax rule for the specific value of Q.

In what follows we will prove results under the assumption that **d(x)** has an acceptance region of the form $\left[K_1\sigma, K_2\sigma\right]$ where $K_1 \le K_2$ [5]. The expression $\sup_{-\infty < \theta < +\infty} R(\theta, d)$ will be minimised[6] by optimal choice of **d** considering that **d** is represented by the choice of K_1 and K_2. This means that α_d^* and β_d^* will have to be expressed as functions of K_1 and K_2 .

---

[3] α_d^* is commonly referred to as the **size** of the test **d**.

[4] Note also that $\beta_d^* = 1 - \inf_{\theta \notin H_0} \alpha_d(\theta)$

[5] The class of such tests is essentially complete and every test of this type is admissible. See T. S. Ferguson, Mathematical Statistics: A Decision Theoretic Approach. Academic Press, New York, 1976. Theorems 3 and 4 of page 223 can be applied to this.

[6] In fact as we shall see later, the infimum with respect to K_1 and K_2 is attained.

before seeking the choice of K_1 and K_2 that provide a minimum value for $\sup_{-\infty<\theta<+\infty} R(\theta,d)$. The expressions for α_d^* and β_d^* as functions of K_1 and K_2 are denoted $\alpha^*(K_1,K_2)$ and $\beta^*(K_1,K_2)$ and are derived in Sections 2 and 3 respectively.

1.3 Preview of Findings

The minimization of $\sup_{-\infty<\theta<+\infty} R(\theta,d)$ with respect to K_1 and K_2, leads to a simple solution that is easy to apply, and has a number of interesting properties. The result obtained shows that the optimal choice of K_1 and K_2 always lie on the line $K_1 + K_2 = a + b$. In other words, the acceptance region $[K_1\sigma, K_2\sigma]$ for any minimax decision rule, is always symmetrically placed relative to the composite hypothesis $[a\sigma, b\sigma]$. This is not surprising given the intrinsic symmetry of the problem.

The K_1 and K_2 values for the optimal decision rule are $\left(K_1^*, a+b-K_1^*\right)$ where $K_1^*\left(\frac{Q}{1+Q}\right)$ is the solution of

$$\Phi(K_1-a)+\Phi(K_1-b)=\frac{Q}{1+Q}$$

and Q is the penalty ratio $Q=\dfrac{p_2}{p_1}$ defined earlier[7].

This equation for K_1^* is readily solvable and hence the method is easy to apply given specific values for a, b and Q .

[7] Where Φ is the standardised Gaussian distribution function

$$\Phi(K)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{K} e^{-\frac{1}{2}u^2}du$$

The optimal solution has a number of characteristic properties outlined below.

Hedging Risk: For the optimal **d**, the associated values of α_d^* and β_d^* will always be in the same ratio as p_1 and p_2 i.e. the optimal **d** will satisfy $\alpha_d^* = Q\beta_d^*$. In other words the optimal decision rule is one that balances the two suprema of error probabilities in proportion to their penalty costs. This is proved in Sections 2 and 4 of this report.

Dual Formulations: The decision theory formulation in terms of loss function and the minimax principle (of Section 1) can be compared to a number of alternative formulations which at first sight appear quite different but can be proven to be equivalent. Each of these alternative formulations focuses on choosing the maximum false alarm probability α_d^* as a criterion which the decision maker desires to have fixed at some small value α_0 and adding some other desired feature . One of these alternative formulations looks for **d** having $\alpha_d^* = \alpha_0$ and also having acceptance region that is symmetric about the null hypothesis set $[a\sigma, b\sigma]$ i.e satisfying $K_1 + K_2 = a + b$. It is shown in Section 2.5.1 that these properties are sufficient to identify a minimax rule. Other duality related results are given in Sections 2.5.3.7 and 3.4.

The duality of these two formulations lies in the fact that the solution of this different formulation is simply $\left(K_1^*, a+b-K_1^*\right)$ where $K_1^*(\alpha_0)$ is the solution of $\Phi(K_1 - a) + \Phi(K_1 - b) = \alpha_0$ which as we will see in section 4, is the equation giving the solution of the minimax formulation.

Hence the choice of p_1 and p_2 and the minimax approach is equivalent to choosing a value for α_0 and choosing K_1 and K_2 to be symmetric about the composite null hypothesis. The link between the two formulations is simply the equation $\alpha_0 = \frac{Q}{1+Q}$ or equivalently $\alpha_0 = \frac{p_2}{p_1 + p_2}$.

The minimax formulation, with penalties p_1 and p_2 defining the loss function, shows how this test can be justified as decision theory. The formulation emphasizing the maximum false alarm probability is more useful for safeguards inspectors who may have less difficult in choosing a value for α_0(a maximum false alarm probability they are willing to tolerate) than in imaging values for p_1 and p_2 .

1.4 Guide to the Report

The Sections 2 to 5 of this report contain the following material.

Section 2 derives expressions for $\alpha^*\left(K_1, K_2\right)$ i.e. expressions for the supremum of the false alarm probabilities. It first studies the symmetry properties as well as the boundary values and gradients of $\alpha^*\left(K_1, K_2\right)$ throughout the domain of definition. It goes on to study the properties of the contours and the implicit functions defined by the contours. These properties are needed subsequently for the minimization of the supremum of the risk function $R\left(\theta, d\right)$ and for the proof of the duality theorems.

Section 3 derives expressions for $\beta^*\left(K_1, K_2\right)$ i.e. expressions for the supremum of the non-detection probabilities. It first studies the symmetry properties as well as the boundary values and gradients of $\beta^*\left(K_1, K_2\right)$ throughout the domain of definition. It goes on study the properties of the contours and the implicit functions defined by the contours. The graphical relationship between these implicit functions and the implicit functions derived earlier in the case of $\alpha^*\left(K_1, K_2\right)$ are then studied. These properties are used subsequently for the minimization of the supremum of the risk function $R\left(\theta, d\right)$.

Section 4 uses the results of Sections 2 and 3 to study the properties of the objective function to be minimized with respect to K_1 and K_2 i.e. study the properties of $\sup\left\{ \alpha_d^* , \mathrm{Q}\beta_d^* \right\}$. Section 4 shows that the rules satisfying $\alpha_d^* = \mathrm{Q}\beta_d^*$ are a sufficient set in that any rule which is not in this set cannot be a minimum point of $\sup\left\{ \alpha_d^* , \mathrm{Q}\beta_d^* \right\}$. Minimisation over the set[8] defined by $\alpha_d^* = \mathrm{Q}\beta_d^*$ is then used to find the optimal decision rule for the specific values of Q , a, b and $\sigma = \sigma_{MUF}$. As mentioned in Section 1.3 the optimal decision rule is described by an acceptance interval $\left[K_1\sigma , K_2\sigma\right]$ symmetric about the null hypothesis $\left[a\sigma, b\sigma\right]$ and characterised by the equation $\Phi\left(K_1 - a\right) + \Phi\left(K_1 - b\right) = \frac{\mathrm{Q}}{1+\mathrm{Q}}$.

Section 5 shows how tolerance limits for hold-up amounts (5.2) or for the effect of uncorrected bias (5.3) can be used to formulate a composite null hypothesis for assessing the material balance. It then goes on in (5.4) to describe the numerical computation of the test acceptance region once the appropriate composite null hypothesis has been established. Finally it compares minimax tests having composite null hypothesis to the tests traditionally used for simple null hypotheses, and looks at numerical examples (5.5). The numerical results also include computations of the power function of minimax tests. Section 5.6 provides a summary of the duality results related to the minimax tests.

Appendix B provides a glossary, section by section, of the mathematical symbol notation used in the report.

We now go on to Section 2 which derives expressions for $\alpha_d\left(\theta\right)$ and α_d^* as well as the properties of these functions.

---

[8] In section 4 of this report, the rules **d** satisfying $\alpha_d^* = \mathrm{Q}\beta_d^*$ are referred to as "gorge rules".

2. Properties of the Rejection Probabilities and Test Size

Section 2 derives expressions for the supremum of the false alarm probabilities i.e. expressions for $\alpha^*(K_1, K_2)$. It shows that $\alpha^*(K_1, K_2)$ symmetric about $K_1 + K_2 = a + b$ and studies the values and gradients of $\alpha^*(K_1, K_2)$ throughout the domain $K_2 \geq K_1$. It goes on study the properties of the contours and the implicit functions defined by the contours. These properties are needed subsequently for the minimization of the supremum of the risk function $R(\theta, d)$. Most of the properties of the contours are summarized in the graph in Figure 2 (see Section 2.5.3.6).

The final subsection (2.5.3.7) shows that among tests lying on the contour $\alpha^*(K_1, K_2)$, the symmetric test is that which has minimum value of $K_2 - K_1$. This result is taken up later in section 2.5.3.8 in discussing the power properties of the symmetric test.

2.1 Properties of the Rejection Probability $\alpha(\theta, K_1, K_2)$

Confining attention to decision rules $\mathbf{d}$ whose acceptance region is an interval denoted $[K_1\sigma, K_2\sigma]$, the notations $\alpha(\theta, K_1, K_2)$ and $\alpha^*(K_1, K_2)$ will from now on be used in place of $\alpha_d(\theta)$ and α_d^* respectively. Note also that both $\alpha(\theta, K_1, K_2)$ and $\alpha^*(K_1, K_2)$ are functions of σ and also that $\alpha^*(K_1, K_2)$ is a function of a and b.

Using the fact that $\alpha_d(\theta)$ denotes $P(d(x) = a_1 | \theta)$ we have that

$$\alpha(\theta, K_1, K_2) \;=\; 2 - \Phi\left(-K_1 + \frac{\theta}{\sigma}\right) - \Phi\left(K_2 - \frac{\theta}{\sigma}\right)$$

where

$$\Phi(K) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{K} e^{-\frac{1}{2}u^2} du$$

Note that this formula defines $\alpha(\theta, K_1, K_2)$ for all values of θ $-\infty < \theta < +\infty$. It provides the false alarm probability for $\theta \in H_0$ and the power for $\theta \notin H_0$.

When $K_1 = K_2$ we have $\alpha(\theta, K_1, K_2) = 1$; $\forall\theta: -\infty < \theta < +\infty$.

Note that $K_1 = K_2$ defines a decision rule which always rejects the null hypothesis, (i.e. rejects with probability 1).

Properties of the Graph of $\alpha(\theta, K_1, K_2)$

Considering $\alpha(\theta, K_1, K_2)$ over the full range of θ (for K_1 and K_2 fixed), it has a very simple graph. $\alpha(\theta, K_1, K_2)$ has a minimum value at θ_{min} where $\theta_{min} = \frac{\sigma}{2}(K_1 + K_2)$. This is easy to verify by solving $\frac{\partial \alpha(\theta, K_1, K_2)}{\partial\theta} = 0$ to find θ_{min} and then verifying that $\frac{\partial^2 \alpha(\theta_{min}, K_1, K_2)}{\partial\theta^2} > 0$.

For $\theta > \theta_{min}$, $\alpha(\theta, K_1, K_2)$ is monotone increasing i.e. $\frac{\partial\alpha}{\partial\theta} > 0$, and tends asymptotically to unity as $\theta \to +\infty$.

For $\theta < \theta_{min}$, $\alpha(\theta, K_1, K_2)$ is monotone decreasing i.e. $\frac{\partial\alpha}{\partial\theta} < 0$, to the minimum at θ_{min} . It tends asymptotically to unity as $\theta \to -\infty$.

Symmetry of $\alpha\left(\theta, K_1, K_2\right)$ about θ_{min}

To show that the graph $\alpha\left(\theta, K_1, K_2\right)$ is symmetric about θ_{min}, it is sufficient to show that,

$$\alpha\left(\theta_{min}+\varepsilon, K_1, K_2\right) = \alpha\left(\theta_{min}-\varepsilon, K_1, K_2\right), \forall \varepsilon > 0,$$

Using the earlier formulae for θ_{min} and $\alpha\left(\theta, K_1, K_2\right)$ we have,

$$\alpha\left(\theta_{min}+\varepsilon, K_1, K_2\right) = 2-\Phi\left(\frac{K_2-K_1}{2}+\frac{\varepsilon}{\sigma}\right)-\Phi\left(\frac{K_2-K_1}{2}-\frac{\varepsilon}{\sigma}\right)$$

From which it is obvious that,

$$\alpha\left(\theta_{min}+\varepsilon, K_1, K_2\right) = \alpha\left(\theta_{min}-\varepsilon, K_1, K_2\right) \quad \text{QED.}$$

Hence the graph $\alpha\left(\theta, K_1, K_2\right)$ is symmetric about $\theta_{min} = \frac{\sigma}{2}\left(K_1 + K_2\right)$.

Some example graphs of $\alpha\left(\theta, K_1, K_2\right)$

Five example graphs of $\alpha\left(\theta, K_1, K_2\right)$ are shown in the Figures 1a -1e below. They illustrate different situations regarding the position of $\theta_{min} = \frac{\sigma}{2}\left(K_1 + K_2\right)$ relative to the null hypthesis interval $\left[a\sigma, b\sigma\right]$. These different situations are important for the properties of $\alpha^*\left(K_1, K_2\right)$ and $\beta^*\left(K_1, K_2\right)$.

For what follows in this section (Section 2), the important distinction is whether $\theta_{min} \le \frac{\sigma}{2}\left(a+b\right)$ (referred to as Case I) or whether $\theta_{min} \ge \frac{\sigma}{2}\left(a+b\right)$ (referred to as Case II). As we will see later, $\alpha^*\left(K_1, K_2\right)$ is represented by different formulae in these two cases. The fact that these two cases are defined with boundary overlap will be treated later.

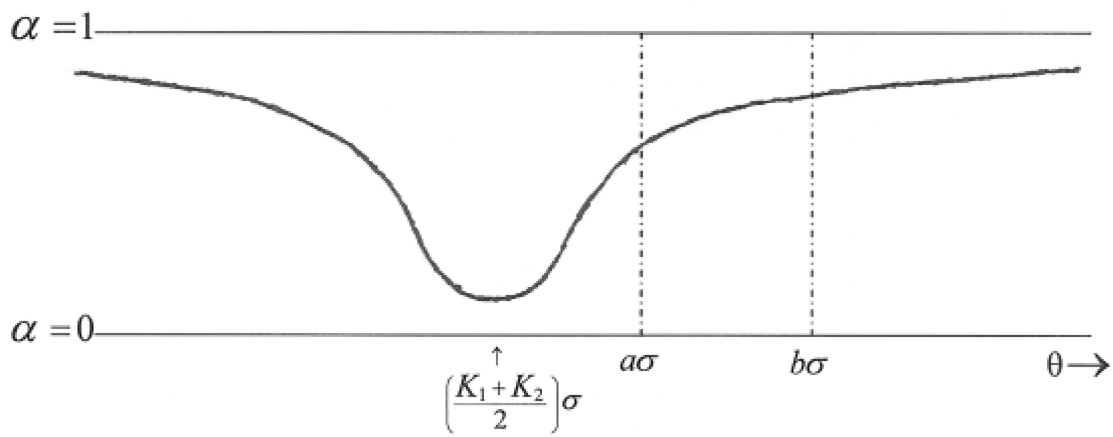


Figure 1.a: Graph of $\alpha\left(\theta, \mathrm{K}_1, \mathrm{K}_2\right)$ as function of θ Case I

Figure 1.a above having $K_1 + K_2 \leq 2a$ and Figure 1.b below having $K_1 + K_2 \geq 2b$, are examples of Case I and Case II respectively.

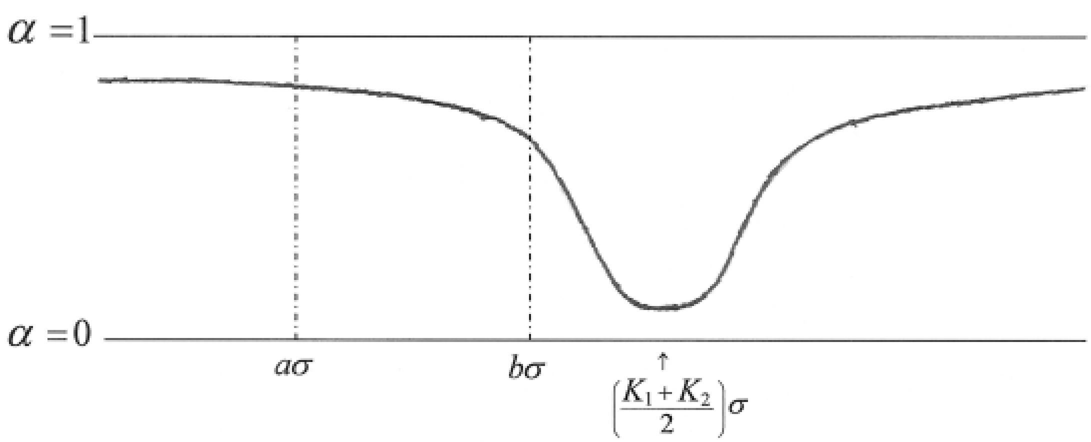


Figure 1.b: Graph of $\alpha\left(\theta, \mathrm{K}_1, \mathrm{K}_2\right)$ as function of θ Case II

Figure 1.c below illustrates the cases where $\left(\mathrm{K}_1, \mathrm{K}_2\right)$ lies on the boundary $K_1 + K_2 = a + b$ which is the intersection of the two cases as we have defined them.

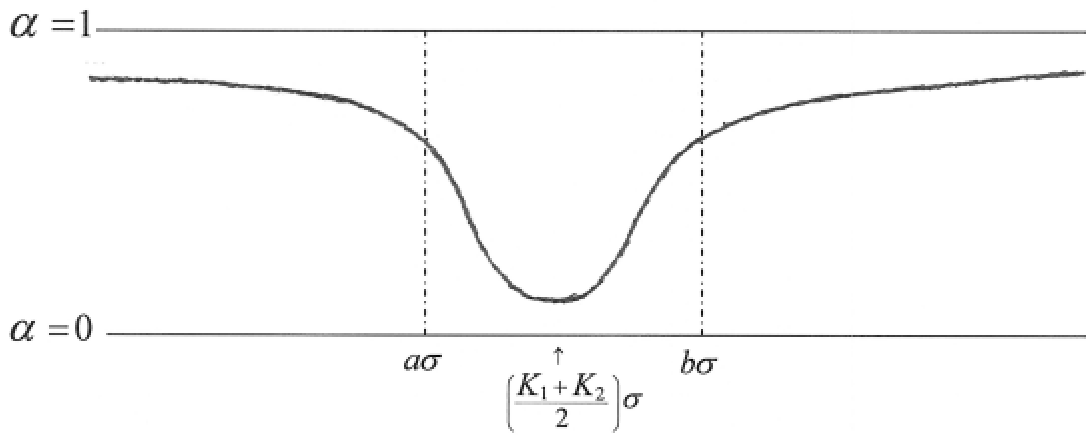


Figure 1.c: Graph of $\alpha\left(\theta, K_1, K_2\right)$ as function of θ $\quad K_1 + K_2 = a + b$

Figure 1.d below, having $2a \leq K_1 + K_2 \leq a + b$ and Figure 1.e below having $2b \geq K_1 + K_2 \geq a + b$, are further illustrations of Case I and Case II.

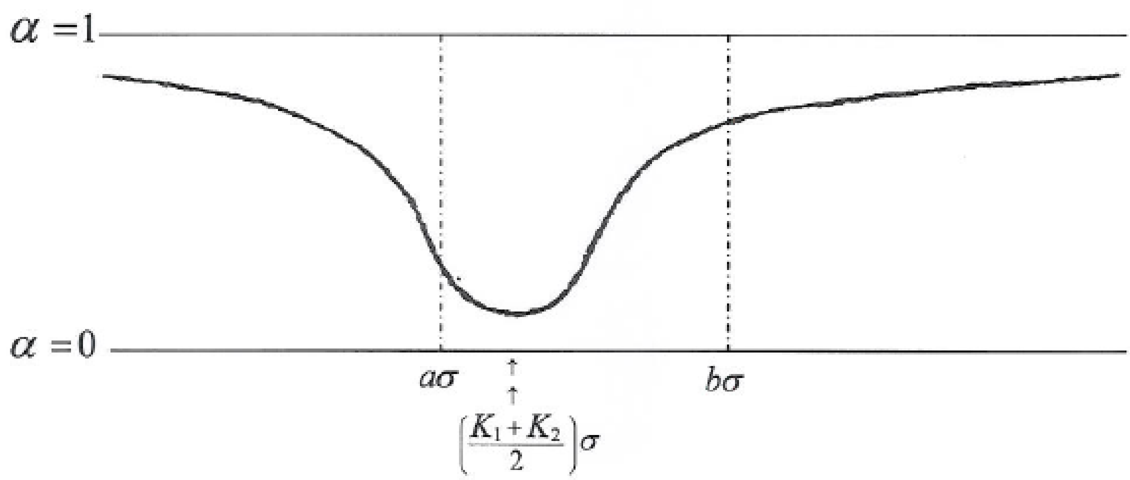


Figure 1.d: Graph of $\alpha\left(\theta, K_1, K_2\right)$ as function of θ Case I

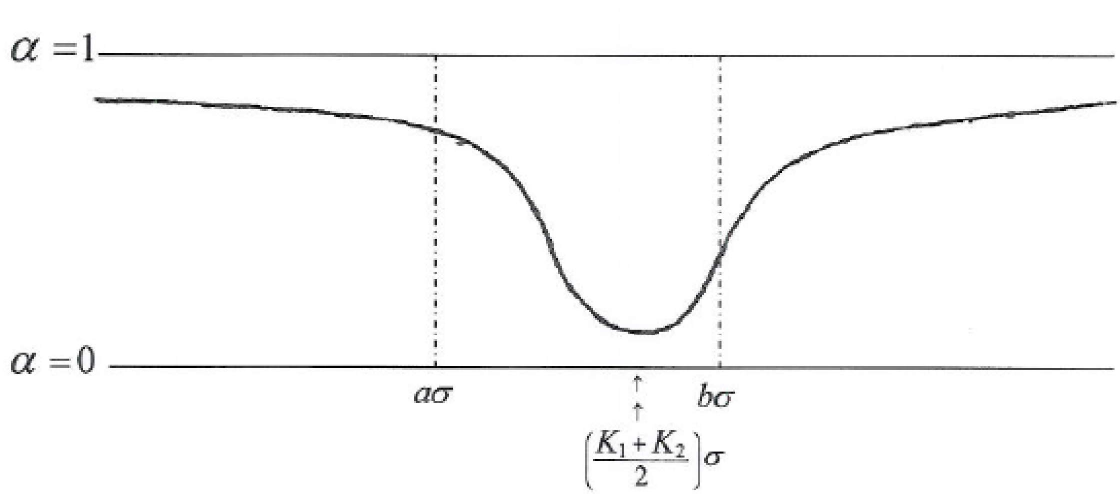


Figure 1.e: Graph of $\alpha\left(\theta, \mathrm{K}_1, \mathrm{K}_2\right)$ as function of θ Case II

Note: For the properties of $\alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$, the distinction between the two subtypes of Case I (e.g. between 1.a and 1.d) is irrelevant, as is also the distinction between the two subtypes of Case II (e.g. between 1.b and 1.e). In Section 3 however, we study $\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$ and the distinction between such subcases is important as they correspond to four different regions of values for $\left(\mathrm{K}_1, \mathrm{K}_2\right)$ having separate formulae for $\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$.

2.2 The Test Size $\alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$

Finding the Expressions for $\alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$

The probability of rejection of H_0 i.e. $\alpha\left(\theta, \mathrm{K}_1, \mathrm{K}_2\right)$ is defined for all values of θ , $-\infty < \theta < +\infty$. $\alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$ is defined by taking sup over the subset of values $\theta \in \left[a\sigma, b\sigma\right]$ i.e.

$$\alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right) = \sup_{\theta \in H_0} \alpha\left(\theta, \mathrm{K}_1, \mathrm{K}_2\right)$$

The determination of the sup depends on the position of $\theta_{\min}$ relative to the interval $\left[a\sigma, b\sigma\right]$. For maximisation, there are two cases depending on whether $\theta_{\min}$ is nearer to $a\sigma$ (Case I) or nearer to $b\sigma$ (Case II). Note $\theta_{\min} = \frac{\sigma}{2}(K_1 + K_2)$.

Case I $\theta_{\min} \leq \frac{\sigma}{2}(a + b)$ i.e. $K_1 + K_2 \leq a + b$; Because $\alpha\left(\theta, K_1, K_2\right)$ is symmetric about $\theta_{\min}$ and has the monotone properties described earlier, the value of $\alpha\left(\theta, K_1, K_2\right)$ at $\theta = b\sigma$ is the largest value in the interval $\left[a\sigma, b\sigma\right]$. Hence we have for Case I that,

$$\alpha^*\left(K_1, K_2\right) = \alpha\left(b\sigma, K_1, K_2\right).$$

Case II $\theta_{\min} \geq \frac{\sigma}{2}(a + b)$ i.e. $K_1 + K_2 \geq a + b$; Again because $\alpha\left(\theta, K_1, K_2\right)$ is symmetric about $\theta_{\min}$ and has the monotone properties described earlier, the value of $\alpha\left(\theta, K_1, K_2\right)$ at $\theta = a\sigma$ is the largest value in the interval $\left[a\sigma, b\sigma\right]$. Hence we have for Case II that $\alpha^*\left(K_1, K_2\right) = \alpha\left(a\sigma, K_1, K_2\right)$.

Note also that when $\theta_{\min} = \frac{\sigma}{2}(a + b)$ i.e. when $K_1 + K_2 = a + b$, the symmetry means that $\theta = a\sigma$ gives a value equal to the value at $\theta = b\sigma$. Hence when $\theta_{\min} = \frac{\sigma}{2}(a + b)$ we have,

$$\alpha^*\left(K_1, K_2\right) = \alpha\left(b\sigma, K_1, K_2\right) = \alpha\left(a\sigma, K_1, K_2\right)$$

Using the expressions for $\theta_{\min}$ and $\alpha\left(\theta, K_1, K_2\right)$, we can express these results for $\alpha^*\left(\mathrm{K}_1,\mathrm{K}_2\right)$ as;

Case I : $K_1 + K_2 \le a + b$

$$\alpha^*\left(\mathrm{K}_1,\mathrm{K}_2\right) = \alpha\left(\mathrm{b}\sigma,\mathrm{K}_1,\mathrm{K}_2\right) = 2 - \Phi\left(-K_1 + b\right) - \Phi\left(K_2 - b\right).$$

Case II: $K_1 + K_2 \ge a + b$

$$\alpha^*\left(\mathrm{K}_1,\mathrm{K}_2\right) = \alpha\left(\mathrm{a}\sigma,\mathrm{K}_1,\mathrm{K}_2\right) = 2 - \Phi\left(-K_1 + a\right) - \Phi\left(K_2 - a\right).$$

These expressions describe $\alpha^*\left(\mathrm{K}_1,\mathrm{K}_2\right)$ for a decision rule **d** whose acceptance region is the interval $\left[K_1\sigma, K_2\sigma\right]$ where $K_2 \ge K_1$. The set of such decision rules can be identified with the half-plane of points $\left\{\left(K_1, K_2\right): K_2 \ge K_1\right\}$.

Symmetric Rules

Decision rules $\left[K_1\sigma, K_2\sigma\right]$ where $K_1 + K_2 = a + b$ will be referred to as symmetric rules (They have acceptance regions that are symmetric about the null hypothesis). We noted earlier that for such rules we have,

$$\alpha^*\left(\mathrm{K}_1,\mathrm{K}_2\right) = \alpha\left(\mathrm{b}\sigma,\mathrm{K}_1,\mathrm{K}_2\right) = \alpha\left(\mathrm{a}\sigma,\mathrm{K}_1,\mathrm{K}_2\right)$$

In other words for symmetric rules, the maximum false alarm probability is attained by $\alpha\left(\theta,\mathrm{K}_1,\mathrm{K}_2\right)$ at the two end points of the null hypothesis.

Continuity across the Boundary $K_1 + K_2 = a + b$

It is easy to show that the two expressions (i.e. Case I and Case II), give the same value for $\alpha^*\left(\mathrm{K}_1,\mathrm{K}_2\right)$ along the line $K_1 + K_2 = a + b$. Substituting $K_2 = a + b - K_1$ in both expressions and using $\Phi(-u) = 1 - \Phi(u); \forall u$ gives the result.

Note also that along the line $K_1 + K_2 = a + b$, the value $\alpha^*(\mathrm{K}_1, \mathrm{K}_2)$ can also be written,

$$\alpha^*(\mathrm{K}_1, a+b-\mathrm{K}_1) \quad = 2 - \Phi(-K_1 + b) - \Phi(a - K_1)$$

or equivalently,

$$\alpha^*(a+b-\mathrm{K}_2, \mathrm{K}_2) \quad = 2 - \Phi(K_2 - b) - \Phi(K_2 - a)$$

2.3 Symmetry Properties of $\alpha^*(\mathrm{K}_1, \mathrm{K}_2)$

In this section we show that the size function $\alpha^*(\mathrm{K}_1, \mathrm{K}_2)$ is symmetric about the line $K_1 + K_2 = a + b$.

Definition of symmetry: Symmetry about $K_1 + K_2 = a + b$ is defined by mapping a point $(\mathrm{K}_1, \mathrm{K}_2)$ into a point $(\mathrm{K}_1', \mathrm{K}_2')$ defined by,

$$\mathrm{K}_1' = a + b - K_2$$

and

$$\mathrm{K}_2' = a + b - K_1 .$$

It is straightforward to show that:

1) $\left((\mathrm{K}_1')', (\mathrm{K}_2')'\right) \equiv (\mathrm{K}_1, \mathrm{K}_2)$.

2) If $(\mathrm{K}_1, \mathrm{K}_2)$ lies on the line $K_1 + K_2 = a + b$ then $\mathrm{K}_1' = \mathrm{K}_1$ and $\mathrm{K}_2' = \mathrm{K}_2$.

3) $K_1 + K_2 > a + b$ if and only if $\mathrm{K}_1' + \mathrm{K}_2' < a + b$

4) The line joining $\left(\mathrm{K}_1, \mathrm{K}_2\right)$ to $\left(\mathrm{K}_1', \mathrm{K}_2'\right)$ is perpendicular to the line $K_1 + K_2 = a + b$ and the points are equidistant from this line.

We can now show that $\alpha^*\left(\mathrm{K}_1', \mathrm{K}_2'\right) = \alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$

Proving that $\alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$ is symmetric about the line $K_1 + K_2 = a + b$

Proof: From 2) above, the result is trivial if $K_1 + K_2 = a + b$. Consider now the case when $K_1 + K_2 > a + b$ and hence by 3) above, $\mathrm{K}_1' + \mathrm{K}_2' < a + b$. By definition when $\mathrm{K}_1' + \mathrm{K}_2' < a + b$ we are Case I and $\alpha^*\left(\mathrm{K}_1', \mathrm{K}_2'\right)$ is given by the Case I formula,

$$\alpha^*\left(\mathrm{K}_1', \mathrm{K}_2'\right) = 2 - \Phi\left(-\mathrm{K}_1' + b\right) - \Phi\left(\mathrm{K}_2' - b\right).$$

Substituting $\mathrm{K}_1' = a + b - K_2$ and $\mathrm{K}_2' = a + b - K_1$, gives

$$\alpha^*\left(\mathrm{K}_1', \mathrm{K}_2'\right) = 2 - \Phi\left(a - \mathrm{K}_1\right) - \Phi\left(\mathrm{K}_2 - a\right)$$

But this however is just the expression derived earlier for $\alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$ when $K_1 + K_2 > a + b$ (Case II). Hence we have shown that when $K_1 + K_2 > a + b$, $\alpha^*\left(\mathrm{K}_1', \mathrm{K}_2'\right) = \alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$. The proof when

$K_1 + K_2 < a + b$, is analogous. Hence $\alpha^*\left(K_1, K_2\right)$ is symmetric about the line $K_1 + K_2 = a + b$. **QED.**

Note that $\alpha\left(\theta, K_1, K_2\right)$ **is NOT Symmetric about** $K_1 + K_2 = a + b$

When we make the transformation: $K_1' = a + b - K_2$ and $K_2' = a + b - K_1$ as before, $\alpha\left(\theta, K_1, K_2\right)$ is not symmetric about $K_1 + K_2 = a + b$. If however we add the additional transformation $\theta' = \left(a + b\right)\sigma - \theta$, we have that $\alpha\left(\theta', K_1', K_2'\right) = \alpha\left(\theta, K_1, K_2\right)$.

2.4 Behaviour of $\alpha^*\left(K_1, K_2\right)$ throughout its Domain

Because of the symmetry of $\alpha^*\left(K_1, K_2\right)$ about $K_1 + K_2 = a + b$ and the definition of K_1 and K_2 whereby we are interested only in the region $K_2 \geq K_1$, we describe here the behaviour of $\alpha^*\left(K_1, K_2\right)$ in the quadrant region defined by $K_1 + K_2 \geq a + b$ and $K_2 \geq K_1$. The boundaries of this quadrant are $K_1 + K_2 = a + b$ and $K_1 = K_2$ and these lines meet at $\left(\frac{a+b}{2}, \frac{a+b}{2}\right)$. The projected values of K_1 and K_2 corresponding to the region are $-\infty < K_1 < +\infty$ and $K_2 \geq \dfrac{a+b}{2}$.

The behaviour of $\alpha^*\left(K_1, K_2\right)$ in this region is determined by the Case II formula,

$$\alpha^*\left(K_1, K_2\right) = 2 - \Phi\left(-K_1 + a\right) - \Phi\left(K_2 - a\right).$$

$\alpha^*\left(K_1, K_2\right)$ has the following properties;

Value of $\alpha^*(K_1, K_2)$ along $K_2 = K_1$

Along the line $K_2 = K_1$ where $K_1 \geq \frac{a+b}{2}$ we have by definition

$$\alpha^*(K_1, K_1) = 2 - \Phi(-K_1 + a) - \Phi(K_1 - a)$$

Since $\Phi(-u) = 1 - \Phi(u); \forall u$ this gives,

$$\alpha^*(K_1, K_1) = 1 ; \forall K_1$$

Value of $\alpha^*(K_1, K_2)$ along $K_1 + K_2 = a + b$

As we have seen earlier (section 2.2), along the line $K_1 + K_2 \geq a + b$, the value $\alpha^*(K_1, K_2)$ can also be written,

$$\alpha^*(K_1, a+b-K_1) = 2 - \Phi(-K_1 + b) - \Phi(a - K_1)$$

or equivalently,

$$\alpha^*(a+b-K_2, K_2) = 2 - \Phi(K_2 - b) - \Phi(K_2 - a)$$

This means that as we move along the line $K_1 + K_2 = a + b$ in the direction $K_2 \to \infty$ and $K_1 \to -\infty$, the value of $\alpha^*(a+b-K_2, K_2)$ is strictly decreasing and $\alpha^*(a+b-K_2, K_2) \to 0$. It runs from a maximal value equal to 1 at $K_1 = \frac{a+b}{2}$ to the value zero at infinity.

The Partial Derivatives of $\alpha^*(\mathrm{K}_1,\mathrm{K}_2)$

Starting from $\alpha^*(\mathrm{K}_1,\mathrm{K}_2) = 2-\Phi(-K_1+a)-\Phi(K_2-a)$ we can find the partial derivatives,

$$\frac{\partial\alpha^*}{\partial \mathrm{K}_1} = \Phi'(\mathrm{K}_1-a) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(\mathrm{K}_1-a)^2} > 0;\quad \forall\, \mathrm{K}_1,\mathrm{K}_2$$

and

$$\frac{\partial\alpha^*}{\partial \mathrm{K}_2} = -\Phi'(\mathrm{K}_2-a) = \frac{-1}{\sqrt{2\pi}}e^{-\frac{1}{2}(\mathrm{K}_2-a)^2} < 0;\quad \forall\, \mathrm{K}_1,\mathrm{K}_2$$

Limits of $\alpha^*(\mathrm{K}_1,\mathrm{K}_2)$ along lines going to Infinity

If we take any point (x,y) i.e. $K_1 = \mathrm{x}$ and $K_2 = \mathrm{y}$ in the Case II region and move along a line from (x,y) to infinity **while staying in the region**, the behaviour of $\alpha^*(\mathrm{K}_1,\mathrm{K}_2) = 2-\Phi(-K_1+a)-\Phi(K_2-a)$ will depend on the direction of movement.

Consider first the case where K_1 is fixed i.e. $K_1 = \mathrm{x}$ and $K_2 \to \infty$. Here $\alpha^*(\mathrm{K}_1,\mathrm{K}_2)$ is monotone decreasing and $\alpha^*(\mathrm{K}_1,\mathrm{K}_2) \to \Phi(K_1-a)$.
Secondly when $K_1 \to -\infty$ and $K_2 \to \infty$, $\alpha^*(\mathrm{K}_1,\mathrm{K}_2)$ is again monotone decreasing and $\alpha^*(\mathrm{K}_1,\mathrm{K}_2) \to 0$.
Thirdly when $K_1 \to +\infty$ and $K_2 \to \infty$, we have $\alpha^*(\mathrm{K}_1,\mathrm{K}_2) \to 1$.

2.5 Properties of the Contours of $\alpha^*(\mathrm{K}_1,\mathrm{K}_2)$

Since $K_2 \geq K_1$ we have $\Phi(K_2-a) \geq \Phi(K_1-a)$ and since

$$\alpha^*(\mathrm{K}_1,\mathrm{K}_2) = 1+\Phi(K_1-a)-\Phi(K_2-a),$$

we have $0 < \alpha^*(\mathrm{K}_1,\mathrm{K}_2) \leq 1$.

For any α_0 where $0 < \alpha_0 \le 1$ the α_0 contour is defined as $\{(\mathrm{K}_1, \mathrm{K}_2), \alpha^*(\mathrm{K}_1, \mathrm{K}_2) = \alpha_0\}$ or equivalently as the set of points $(\mathrm{K}_1, \mathrm{K}_2)$ satisfying $1 + \Phi(K_1 - a) - \Phi(K_2 - a) = \alpha_0$.

The implicit function theorem[9] proves that $1 + \Phi(K_1 - a) - \Phi(K_2 - a) = \alpha_0$ defines K_2 as a single valued function of K_1 for the values K_1 that are possible with the specific value of α_0 [10]. This function will be denoted $\tilde{K}_2(K_1)$. In other words if $(\mathrm{K}_1, \mathrm{K}_2)$ are a feasible solution, the implicit function theorem tells us that K_2 is unique for the given K_1.

We will first show that,

- $\tilde{K}_2(K_1)$ is a strictly monotone increasing function on its domain
- when K_1 is feasible with $\alpha_0 < 1$ then $\tilde{K}_2(K_1) > K_1$ and
- when K_1 is feasible with $\alpha_0 = 1$ then $\tilde{K}_2(K_1) = K_1$.

We will then go on to identify the feasible values of K_1 for each value of α_0.

$\tilde{K}_2(K_1)$ is a strictly monotone increasing function on its domain

Suppose that K_1', K_1'' are feasible points such that $K_1' < K_1''$ and hence $\Phi(K_1' - a) < \Phi(K_1'' - a)$. Since both points are feasible, we have

$$1 + \Phi(K_1' - a) - \Phi(\tilde{K}_2(K_1') - a) = \alpha_0$$

and

$$1 + \Phi(K_1'' - a) - \Phi(\tilde{K}_2(K_1'') - a) = \alpha_0$$

[9] See J. M. H. Olmsted, Real Variables: An Introduction to the Theory of Functions, Appleton-Century-Crofts, Inc., New York 1959. See pages 394, 398 and 419 .

[10] These will be referred to as the feasible values of K_1.

Combining these gives

$$\Phi\left(\tilde{K}_2(K_1'')-a\right)=\Phi\left(\tilde{K}_2(K_1')-a\right)+\Phi\left(K_1''-a\right)-\Phi\left(K_1'-a\right)$$

Since $\Phi\left(K_1'-a\right)<\Phi\left(K_1''-a\right)$ this gives

$$\Phi\left(\tilde{K}_2(K_1'')-a\right)>\Phi\left(\tilde{K}_2(K_1')-a\right)$$

and hence $\tilde{K}_2(K_1'')>\tilde{K}_2(K_1')$ i.e. $\tilde{K}_2(K_1)$ is strictly monotone increasing. QED

The Contour remains in the Feasible Region $K_2 \geq K_1$

When $\alpha_0 < 1$, if $(\mathrm{K}_1, \mathrm{K}_2)$ are a feasible solution of the contour, the point $(\mathrm{K}_1, \mathrm{K}_2)$ remains in the region satisfying $K_2 > K_1$. To see this consider any $(\mathrm{K}_1, \mathrm{K}_2)$ having $1+\Phi\left(K_1-a\right)-\Phi\left(K_2-a\right)=\alpha_0$. This equation combined with the fact that $\alpha_0 < 1$, implies that $\Phi\left(K_1-a\right)<\Phi\left(K_2-a\right)$ and hence that $K_2 > K_1$ as required. It is straightforward to show that $K_2 > K_1$ implies $\alpha_0 < 1$.

When $\alpha_0 = 1$, K_1 being a feasible value for the contour implies $\Phi\left(K_1-a\right)-\Phi\left(K_2-a\right)=0$. Hence when $\alpha_0 = 1$, if $(\mathrm{K}_1, \mathrm{K}_2)$ are a feasible solution, we must have $K_2 = K_1$. Hence $\alpha_0 = 1$ if and only if $K_2 = K_1$.

We will now study the properties of the curve $\tilde{K}_2(K_1)$. This includes identifying the range of feasible values of K_1, the set of feasible pairs $(\mathrm{K}_1, \mathrm{K}_2)$ and properties of the gradient $\dfrac{d\tilde{K}_2}{dK_1}$.

2.5.1 The Point of Intersection of the Contour with $K_1 + K_2 = a + b$

Along the line $K_1 + K_2 = a + b$, the value $\alpha^*(\mathrm{K}_1, \mathrm{K}_2)$ can also be written,

$$\alpha^*(\mathrm{K}_1, a+b-\mathrm{K}_1) = \Phi(K_1 - b) + \Phi(K_1 - a)$$

where $K_1 \le \frac{a+b}{2}$. It is straightforward to show that for $K_1 \le \frac{a+b}{2}$,

- $\alpha^*(\mathrm{K}_1, a+b-\mathrm{K}_1) = 1$ if and only if $K_1 = \frac{a+b}{2}$
- $\alpha^*(\mathrm{K}_1, a+b-\mathrm{K}_1) < 1$ if and only if $K_1 < \frac{a+b}{2}$

At the point $\left(\frac{a+b}{2}, \frac{a+b}{2}\right)$ we have $\alpha^* = 1$.

Moving along the line $K_1 + K_2 = a + b$ with $K_1 \to -\infty$, the value of $\alpha^*(\mathrm{K}_1, a+b-\mathrm{K}_1)$ decreases monotonely and continuously to zero. Hence for every value α_0 where $0 < \alpha_0 \le 1$, there is a unique value K_1^*, $K_1^* \le \frac{a+b}{2}$ such that $\alpha^*\left(K_1^*, a+b-K_1^*\right) = \alpha_0$. In other words, the α_0 contour meets with the line $K_1 + K_2 = a + b$ at $K_1 = K_1^*$ where K_1^* is the solution of $\Phi(K_1 - b) + \Phi(K_1 - a) = \alpha_0$, $0 < \alpha_0 \le 1$. The value K_1^* depends on α_0 and when this is important, the notation $K_1^*(\alpha_0)$ will be used.

Properties of $K_1^*(\alpha_0)$ and Symmetric Rules

For a given α_0 the value of $K_1^*(\alpha_0)$ is found by solving the equation

$$\Phi(K_1 - b) + \Phi(K_1 - a) = \alpha_0$$

Hence for any α_0 , we can find a decision rule $\left(\mathrm{K}_1,\mathrm{K}_2\right)$ having $\alpha^*\left(\mathrm{K}_1,\mathrm{K}_2\right) = \alpha_0$ by solving $\Phi\left(K_1-b\right)+\Phi\left(K_1-a\right) = \alpha_0$ for K_1^* and then setting $K_2^* = a+b-K_1^*$. As we have seen above, for $\alpha_0 < 1$ the value of K_1^* will have $K_1^* < \frac{a+b}{2}$ and hence the value of $K_2^* = a+b-K_1^*$ will have $K_2^* > \frac{a+b}{2}$. For $\alpha_0 = 1$, $K_1^*\left(\alpha_0\right) = \frac{a+b}{2}$.

Note also that we have shown earlier that $\alpha_0 < 1$ implies that all contour points have $\tilde{K}_2\left(K_1\right) > K_1$ and that $\alpha_0 = 1$ implies that all contour points satisfy $\tilde{K}_2\left(K_1\right) = K_1$.

Having $\left(\mathrm{K}_1,\mathrm{K}_2\right)$ lying on the line $K_1 + K_2 = a+b$, defines a decision rule whose acceptance region is symmetric about $\frac{a+b}{2}$. This is because having $K_1 + K_2 = a+b$ is the same as having $K_2 - \frac{a+b}{2} = \frac{a+b}{2} - K_1$. Hence we see that solving the equation $\Phi\left(K_1-b\right)+\Phi\left(K_1-a\right) = \alpha_0$ provides a symmetric decision rule[11] having all its false alarm probabilities $\leq \alpha_0$.

For the simple null hypothesis $b = a$, we have $K_1^*\left(\alpha_0\right) = a + \Phi^{-1}\left(\frac{\alpha_0}{2}\right)$.

For $b = a$ and $\alpha_0 < 1$ this gives $K_1^*\left(\alpha_0\right) < a$.

---

[11] At first sight this has nothing to do with the minimax formulation of Section 1 which motivated us to find an expression for $\alpha^*\left(\mathrm{K}_1,\mathrm{K}_2\right)$. Later we shall show that it has a very fundamental connection.

How does width of symmetric acceptance region vary with $b-a$

The width of the acceptance region is given by $K_2 - K_1$ which in the case of a symmetric rule becomes $a+b-2K_1^*\left(\alpha_0\right)$. To study how $K_2 - K_1$ varies with $b-a$ for a fixed value of α_0, we define $\tau = b-a$ and study how $K_2 - K_1 = 2a+\tau-2K_1^*\left(\alpha_0\right)$ varies as τ increases. Since we have that $\frac{\partial(K_2-K_1)}{\partial\tau} = 1-2\frac{\partial K_1^*}{\partial\tau}$, we need only examine how $K_1^*\left(\alpha_0\right)$ behaves as τ increases.

We now use the fact that $K_1^*\left(\alpha_0\right)$ is the solution of $\Phi\left(K_1-a-\tau\right)+\Phi\left(K_1-a\right) = \alpha_0$ and from the implicit function theorem we have,

$$\frac{\partial K_1^*}{\partial\tau} = \frac{\Phi'\left(K_1^*-a-\tau\right)}{\Phi'\left(K_1^*-a-\tau\right)+\Phi'\left(K_1^*-a\right)} > 0$$

and hence $K_1^*\left(\alpha_0\right)$ is monotone increasing as τ increases. In other words, $K_1^*\left(\alpha_0\right)$ increases as b increases with α_0 and a fixed .

Now $\frac{\partial K_1^*}{\partial\tau}$ can also be written,

$$\frac{\partial K_1^*}{\partial\tau} = \frac{1}{1+e^{\tau\left(\frac{\tau}{2}-K_1^*+a\right)}} \quad .$$

Remembering that $\tau = b - a$, this can also be written,

$$\frac{\partial K_1^*}{\partial \tau} = \frac{1}{1 + e^{(b-a)\left(\frac{a+b}{2} - K_1^*\right)}}$$

and this shows that,

$$\alpha_0 = 1 \text{ implies } \frac{\partial K_1^*}{\partial \tau} = \frac{1}{2}$$

and

$$\alpha_0 < 1 \text{ implies } \frac{\partial K_1^*}{\partial \tau} < \frac{1}{2} \quad .$$

Hence for fixed $\alpha_0 < 1$, $K_2 - K_1$ of the symmetric rule increases as $b - a$ increases. For $\alpha_0 = 1$, $K_2 - K_1$ remains zero for all values of b.

As $b \to +\infty$, $\alpha_0 = \Phi\left(K_1^* - b\right) + \Phi\left(K_1^* - a\right)$ tends to the equation $\alpha_0 = \Phi\left(K_1^* - a\right)$ and hence the solution $K_1^*(\alpha_0, a, b)$ will tend to $\lim_{b\to\infty} K_1^*(\alpha_0, a, b) = a + \Phi^{-1}(\alpha_0)$ and the solution $K_2 = a + b - K_1^*$ will tend to $K_2 = b - \Phi^{-1}(\alpha_0)$.

Since $K_1^*(\alpha_0, a, b)$ is monotone increasing in b we have $\forall\, b$

$$K_1^*(\alpha_0, a, b) < \lim_{b\to\infty} K_1^*(\alpha_0, a, b) = a + \Phi^{-1}(\alpha_0)$$

Hence for $a \le b < +\infty$ we have that,

$$a + \Phi^{-1}\left(\frac{\alpha_0}{2}\right) \le K_1^*(\alpha_0, a, b) < a + \Phi^{-1}(\alpha_0).$$

where $a + \Phi^{-1}\left(\frac{\alpha_0}{2}\right)$ is the solution for $K_1^*(\alpha_0)$ when $b = a$.

Hence $K_1^*(\alpha_0,a,b)$ is bounded over the range $a \le b < +\infty$.

We also have that $\dfrac{\partial K_1^*(\alpha_0,a,b)}{\partial \tau} \to 0$ as $b \to +\infty$. This of course is obvious from the formula $\dfrac{\partial K_1^*}{\partial \tau} = \dfrac{1}{1+e^{\tau\left(\frac{\tau}{2}-K_1^*+a\right)}}$.

Properties of $K_1^*(\alpha_0)$ when $\alpha_0 < \frac{1}{2}$

From the defining equation for $K_1^*(\alpha_0)$, we have that $\Phi(K_1 - a) < \alpha_0$ which gives $K_1 < a + \Phi^{-1}(\alpha_0)$. Note also that in practical applications, α_0 will be a small probability because it is chosen to limit the false alarm risk. For $\alpha_0 < \frac{1}{2}$ we have $\Phi^{-1}(\alpha_0) < 0$ and hence for $\forall\ \alpha_0 < \frac{1}{2}$, we have $K_1^* < a$ and $K_2^* = a + b - K_1^* > b$. This is true for all a and b.

We have also seen that $\dfrac{\partial K_1^*}{\partial \tau} > 0$ for all a and b.

Hence we have that for fixed a and $\alpha_0 < \frac{1}{2}$, $\left|K_1^*(\alpha_0) - a\right|$ is decreasing as b increases.

How does $K_1^*(\alpha_0)$ vary with α_0

The equation $\Phi(K_1^* - b) + \Phi(K_1^* - a) = \alpha_0$ defines $K_1^*(\alpha_0)$ as an implicit function of α_0. Using the implicit function theorem we have

$$\frac{\partial K_1^*}{\partial \alpha_0} = \frac{1}{\Phi'(K_1^* - a) + \Phi'(K_1^* - b)} > 0$$

Hence as α_0 decreases the value of $K_1^*(\alpha_0)$ also decreases tending to $-\infty$ as $\alpha_0 \to 0$. In addition $K_1^*(\alpha_0)$ tending to $-\infty$ implies $\frac{\partial K_1^*}{\partial \alpha_0} \to +\infty$.

For a given α_0, the value of K_1^* is found by solving the equation

$$\Phi(K_1 - b) + \Phi(K_1 - a) = \alpha_0$$

Note however that this equation can also be interpreted as defining α_0 as an implicit function of K_1^*. The implicitly defined α_0 as a function of K_1^* is simply the inverse function of $K_1^*(\alpha_0)$. Hence $\alpha_0(K_1^*)$ has a gradient,

$$\frac{\partial \alpha_0}{\partial K_1^*} = \left[\frac{\partial K_1^*}{\partial \alpha_0}\right]^{-1} = \Phi'(K_1^* - a) + \Phi'(K_1^* - b).$$

This shows that as $K_1^* \to -\infty$, the corresponding value of $\frac{\partial \alpha_0}{\partial K_1^*} \to 0$.

2.5.2 The Point of Intersection of the Contour with $K_1 + K_2 = 2b$

The behaviour of $\alpha^*(\mathrm{K}_1, \mathrm{K}_2)$ in this region is determined by the Case II formula,

$$\alpha^*(\mathrm{K}_1, \mathrm{K}_2) = 2 - \Phi(-K_1 + a) - \Phi(K_2 - a).$$

Along the line $K_1 + K_2 = 2b$, the value $\alpha^*(\mathrm{K}_1, \mathrm{K}_2)$ can also be written,

$$\alpha^*(\mathrm{K}_1, 2b - \mathrm{K}_1) = \Phi(K_1 - a) + \Phi(K_1 + a - 2b)$$

where the range of K_1 is $K_1 \le b$. At the point (b, b) we have $\alpha^* = 1$.

Moving along the line $K_1 + K_2 = 2b$ with $K_1 \to -\infty$, the value of $\alpha^*(\mathrm{K}_1, 2b - \mathrm{K}_1)$ decreases continuously and monotonely to zero. Hence for every value α_0 where $0 < \alpha_0 \le 1$, there is a unique value K_1^{*b}, $K_1^{*b} \le b$ such that $\alpha^*\left(K_1^{*b}\ ,\ 2b - K_1^{*b}\right) = \alpha_0$.

In other words, the α_0 contour meets $K_1 + K_2 = 2b$ at $K_1 = K_1^{*b}$ where K_1^{*b} is the solution of $\Phi\left(K_1 + a - 2b\right) + \Phi\left(K_1 - a\right) = \alpha_0,\ \ 0 < \alpha_0 \le 1$. The value K_1^{*b} depends on α_0 and when this is important, the notation $K_1^{*b}\left(\alpha_0\right)$ will be used. For $0 < \alpha_0 \le 1$, $K_1^{*b}\left(\alpha_0\right)$ is finite .

When $\alpha_0 = 1$, $K_1^{*b}(1)$ must satisfy $\Phi\left(K_1 + a - 2b\right) + \Phi\left(K_1 - a\right) = 1$. This implies $K_1 + a - 2b = -\left(K_1 - a\right)$ or $K_1 = b$. Hence we have that $K_1^{*b}(1) = b$ if and only if $\alpha_0 = 1$.

When $\alpha_0 < 1$, $K_1^{*b}\left(\alpha_0\right)$ must satisfy $\Phi\left(K_1 + a - 2b\right) + \Phi\left(K_1 - a\right) = \alpha_0$. Now $\Phi\left(K_1 + a - 2b\right) + \Phi\left(K_1 - a\right) < 1$ implies $K_1 + a - 2b < -\left(K_1 - a\right)$ which gives $K_1 < b$. The converse is straightforward. Hence we have that $K_1^{*b}(1) < b$ if and only if $\alpha_0 < 1$.

When $a < b$ it is a property of $K_1^{*b}\left(\alpha_0\right)$ that $K_1^{*b}\left(\alpha_0\right) > K_1^{*}\left(\alpha_0\right)$

If $\alpha_0 = 1$, we have that $K_1^{*b}(1) = b$ and $K_1^{*}(1) = \dfrac{a+b}{2}$ and hence $K_1^{*b}\left(\alpha_0\right) > K_1^{*}\left(\alpha_0\right)$.

If $\alpha_0 < 1$, we have

$$K_1^{*b}\left(\alpha_0\right) \text{ satisfies } \Phi\left(K_1^{*b}+a-2b\right)+\Phi\left(K_1^{*b}-a\right)=\alpha_0$$

and

$$K_1^{*}\left(\alpha_0\right) \text{ satisfies } \Phi\left(K_1^{*}-b\right)+\Phi\left(K_1^{*}-a\right)=\alpha_0 \quad .$$

We require to show that $K_1^{*b}\left(\alpha_0\right) > K_1^{*}\left(\alpha_0\right)$ **.**

Suppose instead that $K_1^{*b}\left(\alpha_0\right) < K_1^{*}\left(\alpha_0\right)$ which would imply that

$$\Phi\left(K_1^{*b}-a\right)<\Phi\left(K_1^{*}-a\right)$$

Using this and the two defining equations would give,

$$\Phi\left(K_1^{*b}+a-2b\right)>\Phi\left(K_1^{*}-b\right)$$

This implies $K_1^{*b}+a-2b > K_1^{*}-b$ which implies $K_1^{*b} > K_1^{*}$ which is a contradiction. Hence $K_1^{*b}\left(\alpha_0\right) < K_1^{*}\left(\alpha_0\right)$ is not possible.

Now $K_1^{*b}\left(\alpha_0\right) = K_1^{*}\left(\alpha_0\right)$ implies $\Phi\left(K_1^{*b}-a\right)=\Phi\left(K_1^{*}-a\right)$. Using this and the two defining equations gives,

$$\Phi\left(K_1^{*b}+a-2b\right)=\Phi\left(K_1^{*}-b\right) .$$

If $a < b$, as we assume in this study, we now have $K_1^{*b} = K_1^{*}+b-a$ which is a contradiction of the assumption $K_1^{*b} = K_1^{*}$. Hence $K_1^{*b} = K_1^{*}$ is impossible when $a<b$.

Hence we have shown that when $a<b$ we must have $K_1^{*b} > K_1^{*}$ **QED.**

Note: It is of course obvious that when $a=b$ we have $K_1^{*b} = K_1^{*}$ because then the two lines $K_1 + K_2 = a+b$ and $K_1 + K_2 = 2b$ are the same line.

It is a property of $K_1^{*b}(\alpha_0)$ that $K_1^{*b}(\alpha_0) < K_1^0(\alpha_0)$.

By definition $K_1^{*b}(\alpha_0)$ satisfies $\Phi\left(K_1^{*b}+a-2b\right)+\Phi\left(K_1^{*b}-a\right)=\alpha_0$.
For $0<\alpha_0\leq 1$, $K_1^{*b}(\alpha_0)$ is finite and hence $\Phi\left(K_1^{*b}+a-2b\right)<\alpha_0$.
We can therefore write $\Phi\left(K_1^{*b}-a\right)=\alpha_0-\Phi\left(K_1^{*b}+a-2b\right)<\alpha_0$
By definition $K_1^0(\alpha_0)$ satisfies $\Phi\left(K_1^0-a\right)=\alpha_0$.
Together these two imply $\Phi\left(K_1^{*b}-a\right)<\Phi\left(K_1^0-a\right)$ and hence we have
$K_1^{*b}(\alpha_0)<K_1^0(\alpha_0)$. **QED.**

2.5.3 The Contour $\alpha^*(\mathrm{K}_1,\mathrm{K}_2)=\alpha_0$ as an Implicit Function

2.5.3.1 The Gradient $\frac{d\tilde{K}_2}{dK_1}$ of the Contour $\alpha^*(\mathrm{K}_1,\mathrm{K}_2)=\alpha_0$

As mentioned earlier, the contour $1+\Phi(K_1-a)-\Phi(K_2-a)=\alpha_0$ defines $\tilde{K}_2(K_1)$ as an implicit function of K_1 for the values K_1 that are possible with the specific value of α_0 (implicit function theorem). Since we are studying the Case II region $K_1+K_2\geq a+b$ and $K_2\geq K_1$, the contour points of interest will be those having $K_1+\tilde{K}_2(K_1)\geq a+b$ and $\tilde{K}_2\geq K_1$.
Note that we have shown earlier that when K_1 is feasible with $\alpha_0<1$, then $\tilde{K}_2(K_1)>K_1$ and when K_1 is feasible with $\alpha_0=1$, then $\tilde{K}_2(K_1)=K_1$.

From the implicit function theorem we have,

$$\frac{d\tilde{K}_2}{dK_1}=\frac{\Phi'(\mathrm{K}_1-a)}{\Phi'(\tilde{\mathrm{K}}_2-a)}=e^{\frac{1}{2}\left(\mathrm{K}_1+\tilde{\mathrm{K}}_2-2a\right)\left(\tilde{\mathrm{K}}_2-\mathrm{K}_1\right)}>0;\ \forall\,\mathrm{K}_1,\tilde{\mathrm{K}}_2$$

where $\Phi'(u)$ denotes $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$.

Note that when $\alpha_0 < 1$ and hence $\tilde{K}_2(K_1) > K_1$ and when $b > a$ and hence $K_1 + \tilde{K}_2(K_1) \geq a + b$ implies $K_1 + \tilde{K}_2(K_1) > 2a$, we have that $\frac{d\tilde{K}_2}{dK_1} > 1$ at any point on the α_0 contour.

The Feasible Values of K_1

The fact that $\tilde{K}_2(K_1)$ is monotone increasing (the gradient is positive for all K_1) suggests that the range of feasible K_1 is a set of values $K_1 \geq K_1^*$. In fact however for any specific $\alpha_0 < 1$ there is an upper limit on the feasible values of K_1. This upper limit is denoted K_1^0 and defined by the equation $\Phi\left(K_1^0 - a\right) = \alpha_0$ i.e. $K_1^0 = a + \Phi^{-1}\left(\alpha_0\right)$. Since the equation $\Phi\left(K_1^* - b\right) + \Phi\left(K_1^* - a\right) = \alpha_0$ defines K_1^* we have that $\Phi\left(K_1^* - a\right) < \Phi\left(K_1^0 - a\right)$. This gives $K_1^* < K_1^0$.

We will now show that the feasible set of values for K_1 is the semi-open interval $\left[K_1^*, K_1^0\right)$. The proof shows **firstly** that for any $K_1 \in \left[K_1^*, K_1^0\right)$, $\exists\, K_2$ such that $\left(K_1, K_2\right)$ lies on the contour and **secondly** that for any $K_1 \geq K_1^0$ there can be no K_2 such that $\left(K_1, K_2\right)$ lies on the contour.

Proof of first part:

We consider any $K_1' \in \left[K_1^*, K_1^0\right)$ and we wish to show $\exists\, K_2$ such that $\left(K_1', K_2\right)$ satisfies $1 + \Phi\left(K_1' - a\right) - \Phi\left(K_2 - a\right) = \alpha_0$. We can write the contour as $\Phi(K_2 - a) = 1 - \left[\alpha_0 - \Phi(K_1' - a)\right]$. Since $K_1' < K_1^0$, we have $\Phi(K_1' - a) < \Phi(K_1^0 - a)$ and since $\Phi(K_1^0 - a) = \alpha_0$ we have $\Phi(K_1' - a) < \alpha_0$. Hence $0 < \alpha_0 - \Phi(K_1' - a) < 1$ and therefore $0 < 1 - \left[\alpha_0 - \Phi(K_1' - a)\right] < 1$.

Hence a finite value $K_2 - a$ exists as solution for the equation $\Phi(K_2 - a) = 1 - \left[\alpha_0 - \Phi(K_1' - a)\right]$. In other words a finite solution K_2 exists for the contour equation. This solution is unique and is therefore the unique point of intersection between the contour and the line $K_1 = K_1'$. QED.

Proof of second part:

We now look at $K_1' \geq K_1^0$ which can be treated as two cases. First consider $K_1' > K_1^0$ which implies that $\Phi(K_1' - a) > \Phi(K_1^0 - a)$ or $\Phi(K_1' - a) > \alpha_0$. Now assume that $\exists\, K_2$ such that (K_1', K_2) satisfies the contour equation in the form $\Phi(K_2 - a) = 1 + \left[\Phi(K_1' - a) - \alpha_0\right]$. Since $\Phi(K_1' - a) > \alpha_0$ there is no solution to the equation. Hence for $K_1' > K_1^0$ the contour has no intersection with the line $K_1 = K_1'$.

Now consider the case $K_1' = K_1^0$ which implies $\Phi(K_1' - a) = \alpha_0$. The contour equation now reduces to $\Phi(K_2 - a) = 1$. Again there is no finite solution $K_2 - a$ to this equation. QED.

Note that we can also show that for $\forall K_1' < K_1^*(\alpha_0)$ the line $K_1 = K_1'$ has no intersection in $K_1 + K_2 \geq a + b$ with $\alpha^*(K_1, K_2) = \alpha_0$. To see this consider any point (K_1', K_2) having $K_2 \geq a + b - K_1'$. By definition,

$$\alpha^*(K_1', K_2) = 1 + \Phi(K_1' - a) - \Phi(K_2 - a)$$

Since $K_2 \geq a + b - K_1'$ and Φ is monotone, we have

$$\alpha^*(K_1', K_2) \leq 1 + \Phi(K_1' - a) - \Phi(b - K_1')$$

i.e. $$\alpha^*(K_1', K_2) \leq \Phi(K_1' - a) + \Phi(K_1' - b).$$

Now $K_1' < K_1^*(\alpha_0)$ implies

$$\Phi(K_1' - a) < \Phi(K_1^* - a) \text{ and } \Phi(K_1' - b) < \Phi(K_1^* - b).$$

Hence we have $\alpha^*(K_1', K_2) < \Phi(K_1^* - a) + \Phi(K_1^* - b) = \alpha_0$

In other words, no point (K_1', K_2) having $K_1' < K_1^*(\alpha_0)$ and $K_2 \geq a + b - K_1'$ can satisfy $\alpha^*(K_1, K_2) = \alpha_0$. QED.

2.5.3.2 The Value of $\tilde{K}_2(K_1)$ is Unbounded on $K_1 \in [K_1^*, K_1^0)$

Proof: Consider that we have a monotone increasing sequence of values $\{K_1^{(n)}, n = 1, 2, 3, \ldots\}$ such that each $K_1^{(n)} \in [K_1^*, K_1^0)$ and $\{K_1^{(n)} \to K_1^0\}$. This implies that $\Phi(K_1^{(n)} - a) \to \Phi(K_1^0 - a) = \alpha_0$

Each $\tilde{K}_2(K_1^{(n)})$ satisfies the contour equation,

$$\Phi\left(\tilde{K}_2\left(K_1^{(n)}\right)-a\right)=1+\Phi\left(K_1^{(n)}-a\right)-\alpha_0$$

This implies that $\Phi\left(\tilde{K}_2\left(K_1^{(n)}\right)-a\right)$ is monotone increasing and that $\Phi\left(\tilde{K}_2\left(K_1^{(n)}\right)-a\right)\to 1$.

Now assume that $\tilde{K}_2\left(K_1\right)$ is bounded on $K_1\in\left[K_1^*,K_1^0\right)$ i.e. $\exists B$ finite and $\tilde{K}_2\left(K_1\right)<B \quad \forall K_1\in\left[K_1^*,K_1^0\right)$. Now since B is finite, $\Phi\left(B-a\right)<1$. Since for each $K_1^{(n)}$ we have $\tilde{K}_2\left(K_1^{(n)}\right)-a<B-a$ we consequently have

$$\Phi\left(\tilde{K}_2\left(K_1^{(n)}\right)-a\right)<\Phi\left(B-a\right)<1 \quad \forall n$$

This implies

$$\lim_{n\to+\infty}\Phi\left(\tilde{K}_2\left(K_1^{(n)}\right)-a\right)\leq\Phi\left(B-a\right)<1 \quad \forall n$$

This is a contradiction and hence $\tilde{K}_2\left(K_1\right)$ cannot be bounded. Hence $\tilde{K}_2\left(K_1\right)$ is unbounded. QED.

Note that we readily have an additional fact which is that for any monotone increasing sequence of values $\left\{K_1^{(n)},n=1,2,3,...\right\}$ such that each $K_1^{(n)}\in\left[K_1^*,K_1^0\right)$ and $\left\{K_1^{(n)}\to K_1^0\right\}$, we have that $\tilde{K}_2\left(K_1^{(n)}\right)\to$ monotonely to $+\infty$. The proof uses the fact (section 2.5) that $\tilde{K}_2\left(K_1\right)$ is strictly monotone increasing.

2.5.3.3 Behaviour of $\frac{d\tilde{K}_2}{dK_1}$ along the line $K_1 + K_2 = a + b$

In section 2.5.3.1 we derived,

$$\frac{d\tilde{K}_2}{dK_1} = e^{\frac{1}{2}\left(K_1+\tilde{K}_2-2a\right)\left(\tilde{K}_2-K_1\right)} > 0;\ \forall K_1 \in \left[K_1^*, K_1^0\right)$$

When $\alpha_0 = 1$, we have $K_1^*\left(\alpha_0\right) = \frac{a+b}{2}$ and therefore $\tilde{K}_2\left(K_1^*\right) = \frac{a+b}{2}$.

Hence $\tilde{K}_2\left(K_1^*\right) = K_1^*$ and we have $\frac{d\tilde{K}_2}{dK_1} = 1$ at the point $\left(\frac{a+b}{2}, \frac{a+b}{2}\right)$.

When $\alpha_0 < 1$, $K_1^*\left(\alpha_0\right) < \frac{a+b}{2}$ and therefore $\tilde{K}_2\left(K_1^*\right) > \frac{a+b}{2}$ whence we have that $\tilde{K}_2\left(K_1^*\right) - K_1^* > 0$.

When $b > a$, $K_1^* + \tilde{K}_2\left(K_1^*\right) = a+b$ implies $K_1^* + \tilde{K}_2\left(K_1^*\right) > 2a$ i.e. when $\alpha_0 < 1$ and $b > a$, both terms in the exponential defining the gradient are greater than zero and hence we have $\frac{d\tilde{K}_2}{dK_1} > 1$ at the point $K_1 = K_1^*\left(\alpha_0\right)$, $\tilde{K}_2\left(K_1^*\right) = a+b-K_1^*$.

When $\alpha_0 < 1$ and $b = a$ we have $K_1^* + \tilde{K}_2\left(K_1^*\right) = 2a$ and hence the power of the exponential is zero giving $\frac{d\tilde{K}_2}{dK_1} = 1$ at the point where the contour meets $K_1 + K_2 = 2a$.

Hence we have shown that along the line $K_1 + K_2 = a + b$, when

$\alpha_0 < 1$ and when $b > a$, we have $\frac{d\tilde{K}_2}{dK_1} > 1$. In any other condition,

we have $\frac{d\tilde{K}_2}{dK_1} = 1$.

Conclusion: In the situations of interest to us ($\alpha_0 < 1$ and $b > a$), the contour curve departs from the line $K_1 + K_2 = a + b$ with an initial gradient greater than 1.

We now go on to identify other properties of this gradient i.e. how it depends on α_0 and how for fixed α_0 it changes as the point on the contour moves away from $K_1 + K_2 = a + b$.

2.5.3.4 The second derivative of $\tilde{K}_2(K_1)$

Again from the implicit function theorem we have,

$$\frac{d^2\tilde{K}_2}{dK_1^2} = \frac{\Phi'(K_1 - a)}{\left[\Phi'(\tilde{K}_2 - a)\right]^2}\left[(\tilde{K}_2 - a)\Phi'(K_1 - a) - (K_1 - a)\Phi'(\tilde{K}_2 - a)\right]$$

From this we can show that $\frac{d^2\tilde{K}_2}{dK_1^2} = 0$ if and only if $\tilde{K}_2(K_1) = K_1$. We have proved earlier that $\tilde{K}_2(K_1) = K_1$ if and only if $\alpha_0 = 1$.

We have also proved earlier that when $\alpha_0 < 1$, $\tilde{K}_2(K_1) > K_1$.

Now when $\tilde{K}_2(K_1) > K_1$ the above formula can be used to show that $\dfrac{d^2\tilde{K}_2}{dK_1^2} > 0$. The proof requires nothing more than the fact that $u\, e^{\frac{1}{2}u^2}$ is a monotone increasing function. The fact that $\dfrac{d^2\tilde{K}_2}{dK_1^2} = 0$ **only if** $\tilde{K}_2(K_1) = K_1$ also depends on this monotone property.

The fact that $\dfrac{d^2\tilde{K}_2}{dK_1^2} > 0$ when $\alpha_0 < 1$ tells us that the gradient $\dfrac{d\tilde{K}_2}{dK_1}$ is monotone increasing. Since the gradient starts at with a value $\dfrac{d\tilde{K}_2}{dK_1} > 1$ at $K_1^*(\alpha_0)$ [when $\alpha_0 < 1$, section 2.5.3.3] and then increases, we have $\dfrac{d\tilde{K}_2}{dK_1} > 1 \quad \forall K_1 \in \left[K_1^*, K_1^0\right)$.

2.5.3.5 The line $K_1 = K_1^0$ is an asymptote to the α_0 contour ($\alpha_0 < 1$).

We have shown earlier when $\alpha_0 < 1$ that,

- the feasible values of K_1 are the bounded set $K_1 \in \left[K_1^*, K_1^0\right)$ (2.5.3.1),
- the gradient $\dfrac{d\tilde{K}_2}{dK_1} > 1$ and is monotone increasing $\forall K_1 \in \left[K_1^*, K_1^0\right)$ (2.5.3.4),
- the contour $\tilde{K}_2(K_1)$ is unbounded on $K_1 \in \left[K_1^*, K_1^0\right)$ and in addition that $\tilde{K}_2(K_1) \rightarrow +\infty$ as $K_1 \rightarrow K_1^0$ (2.5.3.2).

We earlier (2.5.3.1) derived the expression for the gradient as,

$$\frac{d\tilde{K}_2}{dK_1} = \frac{\Phi'\left(K_1 - a\right)}{\Phi'\left(\tilde{K}_2 - a\right)} = e^{\frac{1}{2}\left(K_1 + \tilde{K}_2 - 2a\right)\left(\tilde{K}_2 - K_1\right)}$$

Since K_1 is bounded and $\tilde{K}_2\left(K_1\right) \to +\infty$ as $K_1 \to K_1^0$, we see from this expression that

$$\frac{d\tilde{K}_2}{dK_1} \to +\infty \quad \text{as} \quad K_1 \to K_1^0$$

This shows that the line $K_1 = K_1^0$ is an asymptote to the α_0 contour.

2.5.3.6 The Graph of $\tilde{K}_2\left(K_1\right)$

The results we have proved (Sections 2.1 to Section 2.5.3.5) provide a picture of $\tilde{K}_2\left(K_1\right)$. The graph of the contour when $\alpha_0 < 1$ is shown in Figure 2 below. The half contour in the region $K_1 + K_2 \geq a + b$ is defined for $K_1 \in \left[K_1^*, K_1^0\right)$ as a function starting from the point $\left(K_1^*, a + b - K_1^*\right)$ and for $\alpha_0 < 1$ and $b > a$, having a gradient $\frac{d\tilde{K}_2}{dK_1}$ which is greater than 1 and which is always increasing. The contour (for $\alpha_0 < 1$) remains in the region satisfying $K_2 > K_1$ and has the line $K_1 = K_1^0$ as asymptote. This is the situation shown in Figure 2.

When $\alpha_0 = 1$, the graph is the half line $K_2 = K_1$ having $K_1 \geq \frac{a+b}{2}$.

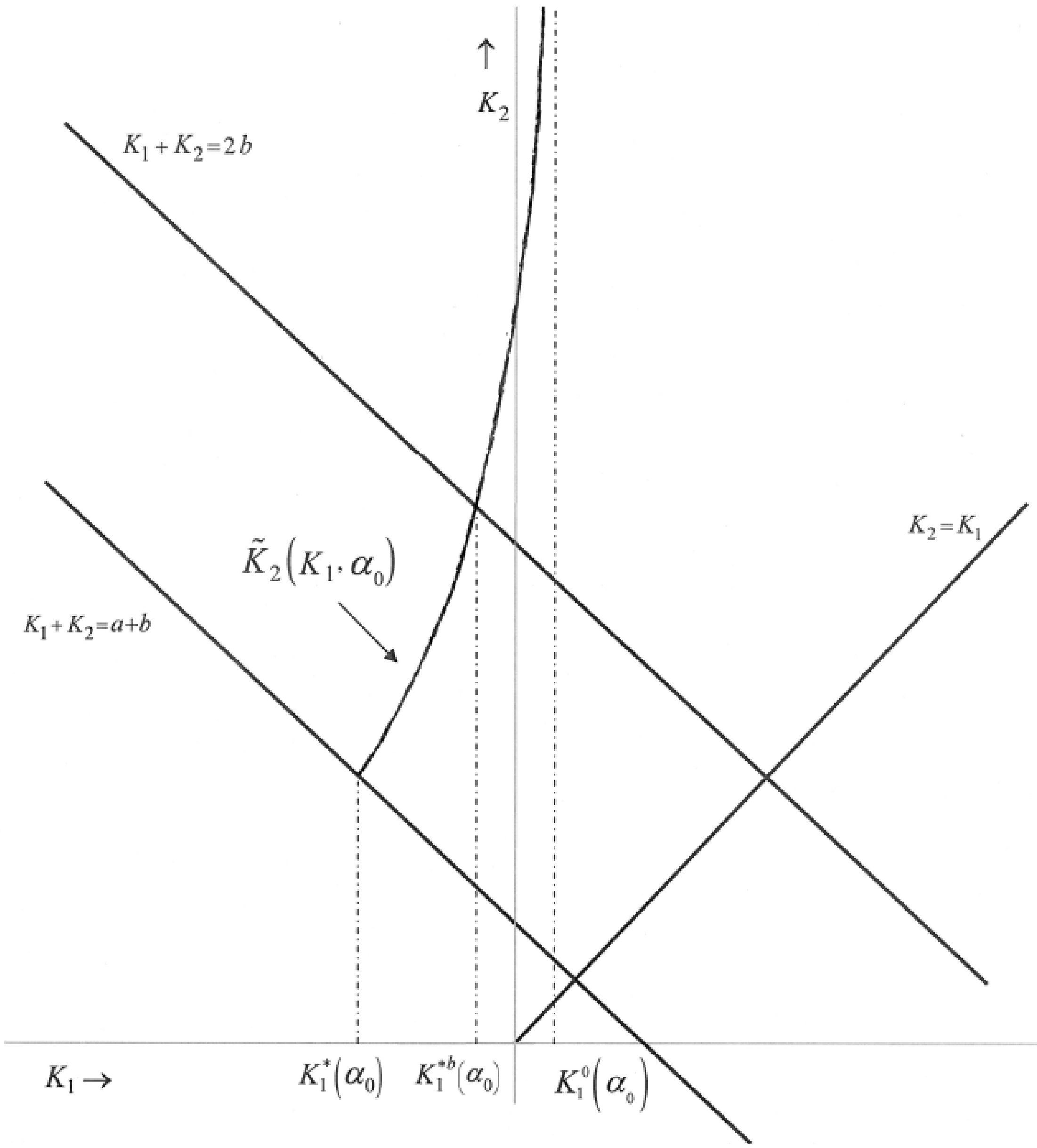


Figure 2: Graph of the contour $\tilde{K}_2(K_1, \alpha_0)$ for $K_1 + K_2 \geq a + b$

2.5.3.7 Minimising $K_2 - K_1$ along $\alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right) = \alpha_0$

Consider points $\left(\mathrm{K}_1, \mathrm{K}_2\right)$ lying on the contour $\alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right) = \alpha_0$ where $\alpha_0 < 1$. These points are of the form $\left(\mathrm{K}_1, \tilde{\mathrm{K}}_2\left(\mathrm{K}_1\right)\right)$. We wish to find among these the points having **minimum value** of $\tilde{K}_2 - K_1$ **.** To do this, we shall show that for any K_1' and K_2'' having $K_1^* \le K_1' < K_1'' < K_1^0$ we have $\tilde{K}_2\left(K_1''\right) - K_1'' > \tilde{K}_2\left(K_1'\right) - K_1'$ **.**

To do this we define $f\left(K_1\right) = \tilde{K}_2\left(K_1\right) - K_1$

$$\frac{df}{dK_1} = \frac{d\tilde{K}_2}{dK_1} - 1 \quad \forall K_1 \in \left[K_1^*, K_1^0\right)$$

We have seen earlier that,

$$\frac{d\tilde{K}_2}{dK_1} = e^{\frac{1}{2}\left(\mathrm{K}_1 + \tilde{\mathrm{K}}_2 - 2a\right)\left(\tilde{\mathrm{K}}_2 - \mathrm{K}_1\right)}$$

This expression was used (when $\alpha_0 < 1$ and $b > a$) to show that $\tilde{K}_2\left(K_1\right)$ starts at $K_1^*\left(\alpha_0\right)$ with a value $\dfrac{d\tilde{K}_2}{dK_1} > 1$ and that this gradient is a monotone increasing function of K_1, $\forall K_1 \in \left[K_1^*, K_1^0\right)$.

Hence $\dfrac{df}{dK_1} = \dfrac{d\tilde{K}_2}{dK_1} - 1$ starts at $K_1^*\left(\alpha_0\right)$ with a value $\dfrac{df}{dK_1} > 0$ $\forall K_1 \in \left[K_1^*, K_1^0\right)$. This implies that $f\left(K_1\right) = \tilde{K}_2\left(K_1\right) - K_1$ is strictly monotone increasing $\forall K_1 \in \left[K_1^*, K_1^0\right)$.

Hence $K_1'' > K_1'$ implies,

$$\tilde{K}_2(K_1'') - K_1'' > \tilde{K}_2(K_1') - K_1' \quad \forall K_1', K_1'' \in \left[K_1^*, K_1^0\right).$$ **QED**

Hence we have shown that among the points (K_1, K_2) on the contour $\alpha^*(K_1, K_2) = \alpha_0$ where $\alpha_0 < 1$, the point having the minimum value of $\tilde{K}_2(K_1) - K_1$ is $\left(K_1^*(\alpha_0), \tilde{K}_2(K_1^*)\right)$ where $\tilde{K}_2(K_1^*) = a + b - K_1^*(\alpha_0)$.

2.5.3.8 The Power Properties of the Symmetric Tests

Each symmetric test is strictly unbiased i.e. for $\forall\, \theta \notin [a\sigma, b\sigma]$, the probability of rejection $\alpha_d(\theta)$ is greater than the size of the test.

PROOF: In general (section 2.1) the probability of rejection $\alpha(\theta, K_1, K_2)$ is given by,

$$\alpha(\theta, K_1, K_2) = 2 - \Phi\left(-K_1 + \frac{\theta}{\sigma}\right) - \Phi\left(K_2 - \frac{\theta}{\sigma}\right)$$

For a symmetric test whose size is α_0, this gives,

$$\alpha\left(\theta, K_1^*(\alpha_0), a+b-K_1^*(\alpha_0)\right)$$
$$= 2 - \Phi\left(-K_1^*(\alpha_0) + \frac{\theta}{\sigma}\right) - \Phi\left(a + b - K_1^*(\alpha_0) - \frac{\theta}{\sigma}\right)$$

In Section 2.2 we saw that the size of a symmetric test is given by

$$\alpha_0 = \alpha^*(K_1, K_2) = \alpha(b\sigma, K_1, K_2) = \alpha(a\sigma, K_1, K_2)$$

Proving that each symmetric test is strictly unbiased, means proving that

$$\alpha\left(\theta, K_1^*(\alpha_0), a+b-K_1^*(\alpha_0)\right) > \alpha_0 \qquad \text{for } \forall\, \theta \notin [a\sigma, b\sigma]$$

This can be proved by showing that $\alpha\left(\theta, K_1^*(\alpha_0), a+b-K_1^*(\alpha_0)\right)$

- when $\theta > b\sigma$, rises monotonely from its minimum value at $\theta = b\sigma$ where, as mentioned above, the value is α_0 and,

- when $\theta < a\sigma$, descends monotonely to its minimum value at $\theta = a\sigma$ where again the value is α_0.

We shall now show that the derivative of $\alpha\left(\theta, K_1^*(\alpha_0), a+b-K_1^*(\alpha_0)\right)$ is always positive in the region $\theta \geq b\sigma$ and always negative in the region $\theta \leq a\sigma$ and this demonstrates the above monotone properties.

The derivative with respect to θ is,

$$\frac{1}{\sigma}\left[\Phi'\left(a+b-K_1^*(\alpha_0)-\frac{\theta}{\sigma}\right)-\Phi'\left(-K_1^*(\alpha_0)+\frac{\theta}{\sigma}\right)\right]$$

where $\Phi'(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$

Writing K_1^* in place of $K_1^*(\alpha_0)$ this expression reduces to

$$\frac{1}{\sqrt{2\pi}\,\sigma^2} e^{-\frac{1}{2}\left(K_1^*-\frac{\theta}{\sigma}\right)^2}\left[e^{-\frac{1}{2}\left(a+b-2K_1^*\right)\left(a+b-\frac{2\theta}{\sigma}\right)} -1\right]$$

For $\alpha_0 < 1$ we have that $K_1^*(\alpha_0) < \frac{a+b}{2}$ and hence $a+b-2K_1^* > 0$.

Whether the derivative is positive or negative depends only on the sign of $a+b-\frac{2\theta}{\sigma}$.

When $\theta > b\sigma$, it is easy to show that $a+b-\frac{2\theta}{\sigma}<0$ and hence that the derivative is positive.

Similarily when $\theta < a\sigma$, it is easy to show that $a+b-\frac{2\theta}{\sigma}>0$ and hence that the derivative is negative.

Hence we have proved that every symmetric test is strictly unbiased. QED.

Alternative Proof of Unbiassedness

Note that the unbiassedness property can also be demonstrated using the result of section 2.1 which described the graph of $\alpha\left(\theta, \mathrm{K}_1, \mathrm{K}_2\right)$ for any rule $\left(\mathrm{K}_1, \mathrm{K}_2\right)$. It was shown there that $\alpha\left(\theta, \mathrm{K}_1, \mathrm{K}_2\right)$ has a minimum value at $\theta_{\min}=\frac{\sigma}{2}\left(K_1+K_2\right)$, is symmetric about $\theta_{\min}$, has $\frac{\partial\alpha}{\partial\theta}<0$ for $\theta<\theta_{\min}$ and $\frac{\partial\alpha}{\partial\theta}>0$ for $\theta>\theta_{\min}$.

In the present case we are dealing with a symmetric rule i.e. a rule having $\theta_{\min}=\frac{\sigma}{2}(a+b)$. In such a case $\alpha\left(\theta, \mathrm{K}_1, \mathrm{K}_2\right)$ is symmetric about $\frac{\sigma}{2}(a+b)$ and has $\frac{\partial\alpha}{\partial\theta}<0$ for $\theta<\frac{\sigma}{2}(a+b)$ and $\frac{\partial\alpha}{\partial\theta}>0$ for $\theta>\frac{\sigma}{2}(a+b)$. This provides an alternative proof that the symmetric test $K_1^*\left(\alpha_0\right), a+b-K_1^*\left(\alpha_0\right)$ is unbiased.

Symmetric Test versus Non-Symmetric Test

It is straightforward to demonstrate the traditional behaviour of the power function in the comparison of a symmetric test against a non-symmetric test of the same size i.e. both having the same value of $\alpha^*\left(\mathrm{K}_1,\mathrm{K}_2\right)$. The non-symmetric test will offer better power on one side of the null hypothesis and there will be a loss of power on the other side of the null hypothesis.

PROOF: Comparing the power function of a symmetric test with a non-symmetric test of the same size, means comparing the symmetric test with another test situated on the same contour[12] $\alpha^*\left(\mathrm{K}_1,\mathrm{K}_2\right)=\alpha_0$. In the region $\mathrm{K}_1+\tilde{\mathrm{K}}_2\left(\mathrm{K}_1\right)\geq a+b$ for example, it is necessary to compare $\alpha\left(\theta, K_1^*\left(\alpha_0\right), a+b-K_1^*\left(\alpha_0\right)\right)$ with $\alpha\left(\theta,\mathrm{K}_1,\tilde{\mathrm{K}}_2\left(\mathrm{K}_1\right)\right)$ [13] as K_1 varies in the set $\left[K_1^*\left(\alpha_0\right), K_1^0\right)$. Here we are comparing the symmetric test with tests whose acceptance region is wider (2.5.3.7) and shifted towards larger values. Such a test will offer reduced power for values of $\theta > b\sigma$ and increased power for $\theta < a\sigma$ (i.e. reduced and increased relative to the symmetric test).

To see this it is sufficient to consider the derivative with respect to K_1 of $\alpha\left(\theta,\mathrm{K}_1,\tilde{\mathrm{K}}_2\left(\mathrm{K}_1\right)\right)$. The derivative $\dfrac{d\alpha}{dK_1}$ is always negative when $\theta > b\sigma$ and is always positive when $\theta < a\sigma$.

---

[12] This is a comparison with the other tests of the same size.

[13] Here $\tilde{\mathrm{K}}_2\left(\mathrm{K}_1\right)$ refers to the contour associated with the value α_0. For correctness it should be written $\tilde{\mathrm{K}}_2\left(\mathrm{K}_1,\alpha_0\right)$ but this notation is too cumbersome.

Properties of the Derivative $\frac{d\alpha}{dK_1}$**:** Differentiating $\alpha\left(\theta, \mathrm{K}_1, \tilde{\mathrm{K}}_2\left(\mathrm{K}_1\right)\right)$ with respect to K_1 and using the fact from Section 2.5.3.1 that when $\mathrm{K}_1 + \mathrm{K}_2 \geq a+b$ we have,

$$\frac{d\tilde{K}_2}{dK_1} = \frac{\Phi'\left(\mathrm{K}_1 - a\right)}{\Phi'\left(\tilde{\mathrm{K}}_2 - a\right)} = e^{\frac{1}{2}\left(\mathrm{K}_1 + \tilde{\mathrm{K}}_2 - 2a\right)\left(\tilde{\mathrm{K}}_2 - \mathrm{K}_1\right)} > 0 \;; \forall\, \mathrm{K}_1, \tilde{\mathrm{K}}_2$$

Using this we get an expression for $\frac{d\alpha}{dK_1}$ as,

$$\frac{d\alpha}{dK_1} = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(K_1 - \frac{\theta}{\sigma}\right)^2}\left[1 - e^{\frac{1}{\sigma}\left(\theta - a\sigma\right)\left(\tilde{\mathrm{K}}_2 - \mathrm{K}_1\right)}\right] \;; \forall\, \mathrm{K}_1, \tilde{\mathrm{K}}_2$$

Looking at the expression it is straightforward to show that $\frac{d\alpha}{dK_1} < 0$ if and only if $\theta > a\sigma$ and that $\frac{d\alpha}{dK_1} > 0$ if and only if $\theta < a\sigma$. These facts are true $\forall\, \mathrm{K}_1 \in \left[K_1^*\left(\alpha_0\right), K_1^0\right)$.

Properties of the Power Function $\alpha\left(\theta, \mathrm{K}_1, \tilde{\mathrm{K}}_2\left(\mathrm{K}_1\right)\right)$**:** Now consider that $\theta \notin H_0$ because $\theta < a\sigma$. In this case we immediately have from the above that $\frac{d\alpha}{dK_1} > 0$. This shows that when $\theta < a\sigma$, α increases monotonely as K_1 increases moving away from $K_1^*\left(\alpha_0\right)$. When $\theta < a\sigma$, the power of the unsymmetric test is always greater than that of the symmetric test.

The other case is that $\theta \notin H_0$ because $\theta > b\sigma$. In this case since $a \le b$, we have $\theta > a\sigma$ and hence $\frac{d\alpha}{dK_1} < 0$. This behaviour of $\frac{d\alpha}{dK_1}$ shows that when $\theta > b\sigma$, α decreases monotonely as K_1 increases moving away from $K_1^*(\alpha_0)$. When $\theta > b\sigma$, the power of the unsymmetric test is always less than that of the symmetric test.

The Comparison of Power when $K_1 + K_2 \ge a + b$

The preceding analysis compares the symmetric test with the tests $\left(K_1, \tilde{K}_2(K_1)\right)$ on the contour $\alpha^*(K_1, K_2) = \alpha_0$ for the part of the contour in the region $K_1 + K_2 \ge a + b$. The analogous analysis for the region $K_1 + K_2 \le a + b$ can be carried out using the formula for $\frac{d\tilde{K}_2}{dK_1}$ derived from the Case I formula (Section 2.2) for $\alpha^*(K_1, K_2) = \alpha_0$. This corresponds to the unsymmetric tests of the same size but having $K_1 < K_1^*(\alpha_0)$. In this case the analysis of $\frac{d\alpha}{dK_1}$ shows that when $\theta > b\sigma$ ($\frac{d\alpha}{dK_1} < 0$), α increases monotonely as K_1 decreases moving away from $K_1^*(\alpha_0)$. Hence when $\theta > b\sigma$, the power of the unsymmetric test is always greater than that of the symmetric test.

Similarly it shows that when $\theta < a\sigma$ ($\frac{d\alpha}{dK_1} > 0$), α decreases monotonely as K_1 decreases moving away from $K_1^*(\alpha_0)$. Hence when $\theta < a\sigma$, the power of the unsymmetric test is always less than that of the symmetric test.

3. The Supremum β_d^* of Non-detection Probabilities

3.1 Finding Expressions for $\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$

In this section we find expressions for β_d^* as functions of K_1 and K_2 and then show that β_d^* is symmetric about the line $K_1 + K_2 = a + b$. We then describe the behaviour of β_d^* throughout the region $K_1 + K_2 \geq a + b$ and $K_2 \geq K_1$. From now on the notation $\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$ will be used in place of β_d^*. In Section 1, β_d^* was defined by

$$\beta_d^* = \sup_{\theta \notin H_0} \left[1 - \alpha_d(\theta)\right].$$

This defining expression can also be written,

$$\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right) = 1 - \inf_{\theta \notin H_0} \alpha\left(\theta, \mathrm{K}_1, \mathrm{K}_2\right)$$

where $\theta \notin H_0$ refers to the union $\{\theta < a\sigma\} \cup \{\theta > b\sigma\}$. Note that $\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$ is a function of σ and of a and b.

In Section 2 we found that $\alpha_d(\theta) = P\left(d(x) = RH_0 | \theta\right)$ was given by

$$\alpha\left(\theta, \mathrm{K}_1, \mathrm{K}_2\right) = 2 - \Phi\left(-K_1 + \frac{\theta}{\sigma}\right) - \Phi\left(K_2 - \frac{\theta}{\sigma}\right)$$

for all values of θ $-\infty < \theta < +\infty$.

In Section 2.1 we also saw that $\alpha\left(\theta, K_1, K_2\right)$ has a minimum value at $\theta_{\min}$ where $\theta_{\min} = \frac{\sigma}{2}(K_1 + K_2)$ and that the graph $\alpha\left(\theta, K_1, K_2\right)$ is symmetric about $\theta_{\min}$. We also saw that as θ moves away from $\theta_{\min}$, $\alpha\left(\theta, K_1, K_2\right)$ increases monotonely to 1 as $\theta \to -\infty$ or as $\theta \to +\infty$.

3.1.1 Expressions for $\beta^*\left(K_1, K_2\right)$

Deriving expressions for $\beta^*\left(K_1, K_2\right)$ requires evaluation of expressions for $\inf\limits_{\theta \notin H_0} \alpha\left(\theta, K_1, K_2\right)$ where $\theta \notin H_0$ means $\theta \in \{\theta < a\sigma\} \cup \{\theta > b\sigma\}$.

The evaluation of the inf depends on the position of $\theta_{\min}$ relative to the values $a\sigma$ and $b\sigma$. There are four cases depending on whether $\theta_{\min} \leq a\sigma$ (Region I) or $\theta_{\min} \geq b\sigma$ (Region IV), and when $\theta_{\min} \in \left[a\sigma, b\sigma\right]$ depending on whether $\theta_{\min}$ is in the lower half, i.e. $a\sigma \leq \theta_{\min} \leq \frac{\sigma}{2}\left(a + b\right)$ (Region II) or in the upper half i.e. $\frac{\sigma}{2}\left(a + b\right) \leq \theta_{\min} \leq b\sigma$ (Region III).

The $\inf\limits_{\theta \notin H_0} \alpha\left(\theta, K_1, K_2\right)$ must be studied separately for each of these four regions. The fact that these regions have been defined with overlap on their boundaries will be treated later.

Region I $\theta_{\min} \leq a\sigma$ or equivalently $K_1 + K_2 \leq 2a$;

Because $\alpha\left(\theta, K_1, K_2\right)$ is symmetric about $\theta_{\min}$ and has the monotone properties described earlier, the value of $\alpha\left(\theta, K_1, K_2\right)$ at $\theta = \theta_{\min}$ is the infimum value in the set $\theta \notin H_0$ i.e. $\theta \in \{\theta < a\sigma\} \cup \{\theta > b\sigma\}$. Note that when $\theta_{\min} = a\sigma$, the inf is not attained by a value in the set $\{\theta < a\sigma\}$ but has a value equal to the value at $\alpha\left(a\sigma, K_1, K_2\right) = \alpha\left(\theta_{\min}, K_1, K_2\right)$.

Hence for Region I we have $\beta^*(K_1, K_2) = 1 - \alpha(\theta_{min}, K_1, K_2)$.

Region II $a\sigma \le \theta_{min} \le \frac{\sigma}{2}(a+b)$ or equivalently $2a \le K_1 + K_2 \le a+b$;
Because $\alpha(\theta, K_1, K_2)$ is symmetric about θ_{min} and has the monotone properties described earlier, the value of $\alpha(\theta, K_1, K_2)$ at $\theta = a\sigma$ is the inf value in the set $\theta \notin H_0$ i.e. $\{\theta < a\sigma\} \cup \{\theta > b\sigma\}$. This is because θ_{min} is nearer to $a\sigma$ than to $b\sigma$. [Again the inf is not attained by a point in the set].
Hence for Region II we have, $\beta^*(K_1, K_2) = 1 - \alpha(a\sigma, K_1, K_2)$.

Region III $\frac{\sigma}{2}(a+b) \le \theta_{min} \le b\sigma$ or equivalently $a+b \le K_1 + K_2 \le 2b$;
Because $\alpha(\theta, K_1, K_2)$ is symmetric about θ_{min} and has the monotone properties described earlier, the value of $\alpha(\theta, K_1, K_2)$ at $\theta = b\sigma$ is the infimum value in the set or $\theta \notin H_0$ i.e. $\{\theta < a\sigma\} \cup \{\theta > b\sigma\}$. This is because θ_{min} is nearer to $b\sigma$ than to $a\sigma$. [Again the infimum is not attained by a point in the set].
Hence for Region III we have $\beta^*(K_1, K_2) = 1 - \alpha(b\sigma, K_1, K_2)$.

Region IV $\theta_{min} \ge b\sigma$ or equivalently $2b \le K_1 + K_2$;
Because $\alpha(\theta, K_1, K_2)$ is symmetric about θ_{min} and has the monotone properties described earlier, the value of $\alpha(\theta, K_1, K_2)$ at $\theta = \theta_{min}$ is the infimum value in the set or $\theta \notin H_0$ i.e. $\{\theta < a\sigma\} \cup \{\theta > b\sigma\}$. Note that when $\theta_{min} = b\sigma$, the inf is not attained by a value in the set $\{\theta > b\sigma\}$ but has a value equal to the value at $\alpha(b\sigma, K_1, K_2) = \alpha(\theta_{min}, K_1, K_2)$.
Hence for Region IV we have $\beta^*(K_1, K_2) = 1 - \alpha(\theta_{min}, K_1, K_2)$.

Substituting $\alpha\left(\theta, K_1, K_2\right)$ in definition of $\beta^*\left(K_1, K_2\right)$

Using the expressions for θ_{min} and $\alpha\left(\theta, K_1, K_2\right)$, we can express the expressions for $\beta^*\left(K_1, K_2\right)$ as;

Region I : $K_1 + K_2 \leq 2a$;

$$\beta^*\left(K_1, K_2\right) = 1 - \alpha\left(\theta_{min}, K_1, K_2\right) = -1 + 2\Phi\left(\frac{K_2 - K_1}{2}\right)$$

Region II: $2a \leq K_1 + K_2 \leq a + b$;

$$\beta^*\left(K_1, K_2\right) = 1 - \alpha\left(a\sigma, K_1, K_2\right) = \Phi\left(K_2 - a\right) - \Phi\left(K_1 - a\right)$$

Region III: $a + b \leq K_1 + K_2 \leq 2b$;

$$\beta^*\left(K_1, K_2\right) = 1 - \alpha\left(b\sigma, K_1, K_2\right) = \Phi\left(K_2 - b\right) - \Phi\left(K_1 - b\right)$$

Region IV: $2b \leq K_1 + K_2$;

$$\beta^*\left(K_1, K_2\right) = 1 - \alpha\left(\theta_{min}, K_1, K_2\right) = -1 + 2\Phi\left(\frac{K_2 - K_1}{2}\right)$$

These expressions describe $\beta^*\left(K_1, K_2\right)$ for any decision rule $\mathbf{d}$ whose acceptance region is the interval $\left[K_1\sigma, K_2\sigma\right]$ where $K_2 \geq K_1$.

Continuity at Boundaries between Regions

The fact that we have defined these four cases in an overlapping way is of no importance because,

- when $K_1 + K_2 = 2a$ ($\theta_{\min} = a\sigma$) the formula for Region I gives the same value as the formula for Region II. This is because when $K_1 + K_2 = 2a$ we have,

$$\alpha\left(\theta_{\min}, \mathrm{K}_1, \mathrm{K}_2\right) = \alpha\left(\mathrm{a}\sigma, \mathrm{K}_1, \mathrm{K}_2\right).$$

- when $K_1 + K_2 = a + b$ ($\theta_{\min} = \frac{a+b}{2}\sigma$) the formula for Region II gives the same value as the formula for Region III. This is because when $K_1 + K_2 = a + b$ we have,

$$\alpha\left(\mathrm{b}\sigma, \mathrm{K}_1, \mathrm{K}_2\right) = \alpha\left(\mathrm{a}\sigma, \mathrm{K}_1, \mathrm{K}_2\right).$$

- when $K_1 + K_2 = 2b$ ($\theta_{\min} = b\sigma$) the formula for Region III gives the same value as the formula for Region IV. This is because when $K_1 + K_2 = 2b$ we have,

$$\alpha\left(\theta_{\min}, \mathrm{K}_1, \mathrm{K}_2\right) = \alpha\left(\mathrm{b}\sigma, \mathrm{K}_1, \mathrm{K}_2\right).$$

This proves that the four expressions for $\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$ define a continuous function across the four regions.

3.1.2 Summarising Results for $\alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$ and $\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$

We have found expressions for $\alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$ and $\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$ for all

$[K_1, K_2]$ where $K_2 \geq K_1$. It is useful to summarise these results in terms of the four regions defined for $\beta^*(\mathrm{K}_1, \mathrm{K}_2)$. This gives,

Region I : $K_1 + K_2 \leq 2a$;

$$\alpha^*(\mathrm{K}_1, \mathrm{K}_2) = 1 + \Phi(K_1 - b) - \Phi(K_2 - b)$$
$$\beta^*(\mathrm{K}_1, \mathrm{K}_2) = -1 + 2\,\Phi\left(\frac{K_2 - K_1}{2}\right)$$

Region II: $2a \leq K_1 + K_2 \leq a + b$;

$$\alpha^*(\mathrm{K}_1, \mathrm{K}_2) = 1 + \Phi(K_1 - b) - \Phi(K_2 - b)$$
$$\beta^*(\mathrm{K}_1, \mathrm{K}_2) = \Phi(K_2 - a) - \Phi(K_1 - a)$$

Region III: $a + b \leq K_1 + K_2 \leq 2b$;

$$\alpha^*(\mathrm{K}_1, \mathrm{K}_2) = 1 + \Phi(K_1 - a) - \Phi(K_2 - a)$$
$$\beta^*(\mathrm{K}_1, \mathrm{K}_2) = \Phi(K_2 - b) - \Phi(K_1 - b)$$

Region IV: $2b \leq K_1 + K_2$;

$$\alpha^*(\mathrm{K}_1, \mathrm{K}_2) = 1 + \Phi(K_1 - a) - \Phi(K_2 - a)$$
$$\beta^*(\mathrm{K}_1, \mathrm{K}_2) = -1 + 2\,\Phi\left(\frac{K_2 - K_1}{2}\right)$$

3.1.3 Special Property along the line $K_1 + K_2 = a + b$

Along the line $K_1 + K_2 = a + b$, $\beta^*(\mathrm{K}_1, \mathrm{K}_2) = 1 - \alpha^*(\mathrm{K}_1, \mathrm{K}_2)$. To prove this take the formulae given above for Region II or Region III i.e. Region III gives ;

$$\alpha^*(\mathrm{K}_1, \mathrm{K}_2) = 1 + \Phi(K_1 - a) - \Phi(K_2 - a) \text{ and}$$

$$\beta^*(\mathrm{K}_1, \mathrm{K}_2) = \Phi(K_2 - b) - \Phi(K_1 - b)$$

Then substitute $K_2 = a + b - K_1$ to get expressions in K_1 valid on the line $K_1 + K_2 = a + b$; These are

$$\alpha^*(\mathrm{K}_1, a + b - K_1) = 1 + \Phi(K_1 - a) - \Phi(b - K_1) \text{ and}$$

$$\beta^*(\mathrm{K}_1, a + b - K_1) = \Phi(a - K_1) - \Phi(K_1 - b)$$

Using $\Phi(-u) = 1 - \Phi(u); \forall u$ then gives,

$$\beta^*(\mathrm{K}_1, a + b - K_1) = 1 - \alpha^*(\mathrm{K}_1, a + b - K_1) \text{ and hence,}$$

$\beta^*(\mathrm{K}_1, \mathrm{K}_2) = 1 - \alpha^*(\mathrm{K}_1, \mathrm{K}_2)$, $\forall\ \mathrm{K}_1, \mathrm{K}_2$ satisfying $K_1 + K_2 = a + b$.

QED.

Hence for any decision rule whose acceptance region is symmetric about $\frac{a+b}{2}$, we have $\beta^*(\mathrm{K}_1, \mathrm{K}_2) = 1 - \alpha^*(\mathrm{K}_1, \mathrm{K}_2)$. Such a rule has $\theta_{\min} = \frac{\sigma}{2}(K_1 + K_2) = \frac{\sigma}{2}(a + b)$. Because of the symmetric property of $\alpha(\theta, \mathrm{K}_1, \mathrm{K}_2)$ both of the points $\theta = a\sigma$ or $\theta = b\sigma$ give the sup for $\theta \in [a\sigma, b\sigma]$ and give the inf for $\theta \in \{\theta < a\sigma\} \cup \{\theta > b\sigma\}$.

3.2 Symmetry Properties of $\beta^*(\mathrm{K}_1,\mathrm{K}_2)$

Earlier we proved that $\alpha^*(\mathrm{K}_1,\mathrm{K}_2)$ was symmetric about the line $K_1+K_2=a+b$. Now we show that $\beta^*(\mathrm{K}_1,\mathrm{K}_2)$ has the same property using the same definition of symmetry i.e. $\beta^*(\mathrm{K}_1,\mathrm{K}_2)$ is symmetric about the line $K_1+K_2=a+b$. The difference is that for $\beta^*(\mathrm{K}_1,\mathrm{K}_2)$ we have to show that,

a) Region I maps into Region IV and vice versa,

b) Region II maps into Region III and vice versa.

Region I maps into Region IV: To show that Region I maps into Region IV and vice versa, and that the value of $\beta^*(\mathrm{K}_1,\mathrm{K}_2)$ is preserved, we need to show that,

$$K_1+K_2\le 2a \text{ if and only if } K_1'+K_2'\ge 2b$$

$$\text{and that } \quad \beta^*(\mathrm{K}_1,\mathrm{K}_2)=\beta^*(\mathrm{K}_1',\mathrm{K}_2').$$

This latter is equivalent to

$$\Phi\left(\frac{K_2-K_1}{2}\right)=\Phi\left(\frac{K_2'-K_1'}{2}\right).$$

Symmetry is defined by $\mathrm{K}_1'=a+b-K_2$ and $\mathrm{K}_2'=a+b-K_1$, from which we immediately have $\mathrm{K}_2'-\mathrm{K}_1'=\mathrm{K}_2-\mathrm{K}_1$ QED. It is also straightforward to show that $K_1+K_2\le 2a \Leftrightarrow K_1'+K_2'\ge 2b$ QED.

Region II maps into Region III: To show that Region II maps into Region III and vice versa, we must show that when $K_1' = a + b - K_2$ and $K_2' = a + b - K_1$, $\beta^*(K_1, K_2) = \Phi(K_2 - a) - \Phi(K_1 - a)$ has the same value as $\beta^*(K_1', K_2') = \Phi(K_2' - b) - \Phi(K_1' - b)$.

Substituting for K_2' and K_1' in $\beta^*(K_1', K_2') = \Phi(K_2' - b) - \Phi(K_1' - b)$ we get,

$$\beta^*(K_1', K_2') = \Phi(a - K_1) - \Phi(a - K_2)$$

Using $\Phi(-u) = 1 - \Phi(u); \forall u$, this gives $\beta^*(K_1', K_2') = \beta^*(K_1, K_2)$.

The converse comes analogously from using the reflection transformation to substitute for K_2 and K_1 in $\beta^*(K_1, K_2) = \Phi(K_2 - a) - \Phi(K_1 - a)$.

Hence we have shown that $\beta^*(K_1, K_2)$ is symmetric about the line $K_1 + K_2 = a + b$. QED. It is also straightforward to show that $2a \le K_1 + K_2 \le a + b \Leftrightarrow a + b \le K_1' + K_2' \le 2b$ QED.

3.3 Behaviour of $\beta^*(K_1, K_2)$ throughout its Domain

Since we know that $\beta^*(K_1, K_2)$ is symmetric about $K_1 + K_2 = a + b$ we will only describe the properties of $\beta^*(K_1, K_2)$ in Regions III and IV.

The Partial Derivatives of $\beta^*(K_1, K_2)$

In Region III: $a + b \le K_1 + K_2 \le 2b$; we have,

$$\beta^*(K_1, K_2) = \Phi(K_2 - b) - \Phi(K_1 - b)$$

And hence

$$\frac{\partial \beta^*}{\partial \mathrm{K}_1} = -\Phi'\left(\mathrm{K}_1 - b\right) < 0\,; \;\; \forall\, \mathrm{K}_1, \mathrm{K}_2$$

and

$$\frac{\partial \beta^*}{\partial \mathrm{K}_2} = \Phi'\left(\mathrm{K}_2 - b\right) > 0\,; \;\; \forall\, \mathrm{K}_1, \mathrm{K}_2$$

In Region IV: $2b \le K_1 + K_2$; we have,

$$\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right) = -1 + 2\,\Phi\left(\frac{K_2 - K_1}{2}\right)$$

and hence,

$$\frac{\partial \beta^*}{\partial \mathrm{K}_1} = -\Phi'\left(\frac{K_2 - K_1}{2}\right) < 0\,; \;\; \forall\, \mathrm{K}_1, \mathrm{K}_2$$

and

$$\frac{\partial \beta^*}{\partial \mathrm{K}_2} = \Phi'\left(\frac{K_2 - K_1}{2}\right) > 0\,; \;\; \forall\, \mathrm{K}_1, \mathrm{K}_2$$

Note that the partial derivatives of $\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$ are opposite in sign to the partial derivatives of $\alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$.

$\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$ along the line $K_1 + K_2 = a + b$

Along the line $K_1 + K_2 = a + b$ the value of $\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$ is given by the region III formula $\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right) = \Phi\left(K_2 - b\right) - \Phi\left(K_1 - b\right)$. Substituting $K_2 = a + b - K_1$ gives ,

$$\beta^*\left(K_1, a+b-K_1\right) = \Phi\left(a-K_1\right) - \Phi\left(K_1-b\right)$$

As K_1 decreases $\beta^*\left(K_1, a+b-K_1\right) = 1-\Phi\left(K_1-a\right)-\Phi\left(K_1-b\right)$ is monotone increasing and as $K_1 \to -\infty$, $\beta^*\left(K_1, a+b-K_1\right) \to 1$.

When $K_1 = \frac{a+b}{2}$ and of course $K_2 = \frac{a+b}{2}$, we have

$$\beta^*\left(\frac{a+b}{2}, \frac{a+b}{2}\right) = \Phi\left(a - \frac{a+b}{2}\right) - \Phi\left(\frac{a+b}{2} - b\right)$$

or

$$\beta^*\left(\frac{a+b}{2}, \frac{a+b}{2}\right) = \Phi\left(\frac{a-b}{2}\right) - \Phi\left(\frac{a-b}{2}\right) = 0$$

$\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right) = 0$ **along the line** $K_2 = K_1$

Along the line $K_2 = K_1$, for $K_1 \geq \frac{a+b}{2}$, we have from the defining formula in Region III,

$$\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right) = \beta^*\left(\mathrm{K}_1, \mathrm{K}_1\right) = \Phi\left(K_1-b\right) - \Phi\left(K_1-b\right) = 0$$

and from the defining formula in Region IV we have,

$$\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right) = \beta^*\left(\mathrm{K}_1, \mathrm{K}_1\right) = -1 + 2\Phi(0) = 0$$

The converse is straightforward.

$\beta^*(K_1, K_2)$ along the line $K_1 + K_2 = 2b$

Along the line $K_1 + K_2 = 2b$ the value of $\beta^*(K_1, K_2)$ is given by the Region III formula $\beta^*(K_1, K_2) = \Phi(K_2 - b) - \Phi(K_1 - b)$. Substituting $K_2 = 2b - K_1$ gives,

$$\beta^*(K_1, 2b - K_1) = \Phi(b - K_1) - \Phi(K_1 - b)$$

i.e. $\beta^*(K_1, 2b - K_1) = 2\Phi(b - K_1) - 1$. Note that $K_1 \le b$.

As $K_1 \to -\infty$, $\Phi(b - K_1) \to 1$ (monotonely increasing) and hence $\beta^*(K_1, 2b - K_1) \to 1$ (monotonely increasing).

Behaviour of $\beta^*(K_1, K_2)$ along other lines going to infinity

If we take any point (x,y) i.e. $K_1 = x$ and $K_2 = y$ in the region defined by $K_1 + K_2 \ge a + b$ and $K_2 \ge K_1$ and move along a line from (x,y) to infinity **while staying in the region** i.e. in Regions III and IV. There are essentially two cases to consider.

First is when $K_1 = x$ and $K_2 = y$ is in **Region III** and the line stays in Region III i.e. $a + b \le K_1 + K_2 \le 2b$ and $K_2 \ge K_1$. The line will be of the form $K_1 + K_2 = \lambda$ where $a + b < \lambda < 2b$. In Region III, $\beta^*(K_1, K_2)$ is given by the formula $\beta^*(K_1, K_2) = \Phi(K_2 - b) - \Phi(K_1 - b)$.

Substituting $K_2 = \lambda - K_1$ gives,

$$\beta^*(K_1, \lambda - K_1) = \Phi(\lambda - K_1 - b) - \Phi(K_1 - b).$$

This is equivalent to $\beta^*(K_1, \lambda - K_1) = 1 - \Phi(K_1 + b - \lambda) - \Phi(K_1 - b)$

and shows that as K_1 decreases, $\beta^*\left(\mathrm{K}_1, \lambda - K_1\right)$ is monotone increasing and as $K_1 \to -\infty$, $\beta^*\left(K_1, \lambda - K_1\right) \to 1$.

The second case is when $K_1 = \mathrm{x}$ and $K_2 = \mathrm{y}$ is in **Region IV** and the line stays in Region IV i.e. $2b \le K_1 + K_2$ and $K_2 \ge K_1$ The behaviour of $\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right) = -1 + 2\Phi\left(\frac{K_2 - K_1}{2}\right)$ will depend on the direction of movement.

Consider first the line where K_1 is fixed i.e. $K_1 = \mathrm{x}$ and $K_2 \to \infty$. Here $\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$ is monotone increasing and

$$\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right) = -1 + 2\Phi\left(\frac{K_2 - K_1}{2}\right) \to 1.$$

Secondly consider any line where $K_1 \to -\infty$ and $K_2 \to +\infty$, $\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$ is again monotone increasing and $\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right) \to 1$.

Thirdly for a line where $K_1 \to +\infty$ and $K_2 \to +\infty$, the behaviour depends on whether K_2 grows more rapidly than K_1 (slope > 1). If K_2 grows more rapidly than K_1 , we have $\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right) \to 1$. If however $K_1 \to +\infty$ and $K_2 \to +\infty$ but movement is along a line parallel to the boundary $K_2 = K_1$ we must have $K_2 = \delta + K_1$ where $\delta > 0$. At every point along $K_2 = \delta + K_1$, we have $\beta^*\left(\mathrm{K}_1, \delta + K_1\right) = -1 + 2\Phi\left(\frac{\delta}{2}\right)$ and the limit as $K_1 \to +\infty$ is also equal to this value.

Summary: These results give a picture of $\beta^*\left(\mathrm{K}_1,\mathrm{K}_2\right)$ as zero along $K_2 = K_1$ and as we move towards infinity (linearly) from any point on this line, $\beta^*\left(\mathrm{K}_1,\mathrm{K}_2\right) \to 1$ except for the case of movement parallel to $K_2 = K_1$ when the limit at infinity is <1.

3.4 Minimising $\beta^*\left(\mathrm{K}_1,\mathrm{K}_2\right)$ along the contour $\alpha^*\left(\mathrm{K}_1,\mathrm{K}_2\right) = \alpha_0$

Along the contour $\alpha^*\left(\mathrm{K}_1,\mathrm{K}_2\right) = \alpha_0$ with $\alpha_0 < 1$, the function $\beta^*\left(\mathrm{K}_1,\mathrm{K}_2\right)$ has the special property that it takes its minimum value along the α_0 contour at the symmetric point $\left(K_1^*, a+b - K_1^*\right)$ where $K_1^* = K_1^*(\alpha_0)$ is the solution of $\Phi\left(K_1 - a\right) + \Phi\left(K_1 - b\right) = \alpha_0$.

PROOF: The contour $\alpha^*\left(\mathrm{K}_1,\mathrm{K}_2\right) = \alpha_0$ defines $\tilde{K}_2\left(K_1\right)$ as an implicit function of K_1 for the values K_1 that are possible with the specific value of α_0. The feasible set of values for K_1 is the semi-open interval $\left[K_1^*\left(\alpha_0\right), K_1^0\left(\alpha_0\right)\right)$. The upper limit K_1^0 is defined by the equation $\Phi\left(K_1^0 - a\right) = \alpha_0$ and is the K_1 value associated with the asymptote to the contour. We have shown earlier (Section 2.5) that when K_1 is feasible with $\alpha_0 < 1$ then $\tilde{K}_2\left(K_1\right) > K_1$. The gradient of the contour $\alpha^*\left(\mathrm{K}_1,\mathrm{K}_2\right) = \alpha_0$ is given by (2.5.3.1),

$$\frac{d\tilde{K}_2}{dK_1} = e^{\frac{1}{2}\left(\mathrm{K}_1+\tilde{\mathrm{K}}_2-2a\right)\left(\tilde{\mathrm{K}}_2-\mathrm{K}_1\right)} > 0;\ \forall\, \mathrm{K}_1,\tilde{\mathrm{K}}_2$$

Note also that when $\alpha_0 < 1$, $\dfrac{d\tilde{K}_2}{dK_1} > 1;\ \forall \mathrm{K}_1$ (Section 2.5.3.1).

Along the contour $\alpha^*(\mathrm{K}_1, \mathrm{K}_2) = \alpha_0$ the value of $\beta^*(\mathrm{K}_1, \mathrm{K}_2)$ is

given by $\beta^*(\mathrm{K}_1, \tilde{\mathrm{K}}_2) = \Phi(\tilde{K}_2 - b) - \Phi(K_1 - b)$ in Region III

and given by $\beta^*(\mathrm{K}_1, \tilde{\mathrm{K}}_2) = -1 + 2\Phi\left(\dfrac{\tilde{K}_2 - K_1}{2}\right)$ in Region IV.

Hence for Region III we have,

$$\frac{d\beta^*(\mathrm{K}_1, \tilde{\mathrm{K}}_2(\mathrm{K}_1))}{d\mathrm{K}_1} = \Phi'(\tilde{\mathrm{K}}_2 - b)\frac{d\tilde{K}_2}{dK_1} - \Phi'(\mathrm{K}_1 - b)$$

and for Region IV we have,

$$\frac{d\beta^*(\mathrm{K}_1, \tilde{\mathrm{K}}_2(\mathrm{K}_1))}{d\mathrm{K}_1} = \Phi'\left(\frac{\tilde{K}_2 - K_1}{2}\right)\left[\frac{d\tilde{K}_2}{dK_1} - 1\right]$$

where $\Phi'(u)$ denotes $\dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$.

Both of these expressions for $\dfrac{d\beta^*(\mathrm{K}_1, \tilde{\mathrm{K}}_2(\mathrm{K}_1))}{d\mathrm{K}_1}$ can be shown to be strictly greater than zero provided $b > a$. In both cases the proof uses the fact that $\tilde{K}_2(K_1) > K_1$ which is true because $\alpha_0 < 1$ (Section 2.5).

To establish the result it is sufficient (in both cases) to substitute the expression for $\frac{d\tilde{K}_2}{dK_1}$ and use the definition $\Phi'(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$. In Region IV the proof also uses the fact that $2b \leq K_1 + K_2$ implies $2a < K_1 + K_2$.

The proof gives $\frac{d\beta^*\left(\mathrm{K}_1, \tilde{\mathrm{K}}_2(\mathrm{K}_1)\right)}{d\mathrm{K}_1} > 0$ for all $K_1 \in \left[K_1^*(\alpha_0), K_1^0(\alpha_0)\right)$.

Hence the minimum value of $\beta^*\left(\mathrm{K}_1, \tilde{\mathrm{K}}_2(\mathrm{K}_1)\right)$ is given by the value $\beta^*\left(K_1^*(\alpha_0), \tilde{\mathrm{K}}_2\left(K_1^*(\alpha_0)\right)\right)$ i.e. the value at the point defining the symmetric test $\left(K_1^*(\alpha_0), a+b-K_1^*(\alpha_0)\right)$.QED.

This result can be written,

$$\beta^*\left(\mathrm{K}_1, \tilde{\mathrm{K}}_2\left(\mathrm{K}_1, (\alpha_0)\right)\right) > \beta^*\left(K_1^*(\alpha_0), a+b-K_1^*(\alpha_0)\right)$$
$$\forall K_1 : K_1^*(\alpha_0) < K_1 < K_1^0(\alpha_0).$$

We proved earlier in Section 3.1.3, that for any symmetric test we have $\beta^*\left(K_1^*(\alpha_0), a+b-K_1^*(\alpha_0)\right) = 1-\alpha_0$. Because the derivative is strictly positive, we have as a consequence that $\beta^*\left(\mathrm{K}_1, \tilde{\mathrm{K}}_2(\mathrm{K}_1)\right) > 1-\alpha_0$ for any point having $K_1^*(\alpha_0) < K_1 < K_1^0(\alpha_0)$.

This result reflects the properties mentioned earlier for the power functions of non-symmetric tests (Section 2.5.3.8). Compared to the symmetric test of the same size, the non-symmetric test has larger power on one side of the null hypothesis and reduced power on the other side of the null hypothesis. Here we see an effect of this in that $\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$ will be greater for any non-symmetric test than for the symmetric test of the same size.

3.5 Properties of the Contours of $\beta^*\left(\mathrm{K}_1,\mathrm{K}_2\right)$

In this section we look at the contours $\beta^*\left(\mathrm{K}_1,\mathrm{K}_2\right) = \beta_0$ in the region defined by $K_1 + K_2 \geq a + b$ and $K_2 \geq K_1$ i.e. in Regions III and IV. We have to look at the contour equation separately in each region.

In Region III, the contour equation becomes $\Phi\left(K_2 - b\right) - \Phi\left(K_1 - b\right) = \beta_0$ and for this equation, we are interested only in solutions that are in Region III.

In Region IV, the contour equation becomes $\beta_0 = -1 + 2\Phi\left(\frac{K_2 - K_1}{2}\right)$ and for this equation, we are interested only in solutions that are in Region IV.

The implicit function theorem (opus cit page 19.) proves that each of these equations (in its respective region), implicitly defines K_2 as a continuous single valued function of K_1 for the values K_1 that are possible with the specific value of β_0 [14].

Since $\beta^*\left(\mathrm{K}_1,\mathrm{K}_2\right)$ is continuous across the boundary (Section 3.1.1), the relevant pieces of contour in the two regions intersect the boundary $K_1 + K_2 = 2b$ at the same point. In both regions, the implicit function will be denoted $\ddot{K}_2\left(K_1\right)$ and the properties of $\ddot{K}_2\left(K_1\right)$ will be deduced from the appropriate equation in each region.

[14] These will be referred to here as the "feasible values of K_1". Saying that K_1' is a feasible value for the Region III contour, is used to mean that $\exists K_2$ such that $\beta^*\left(\mathrm{K}_1',\mathrm{K}_2\right) = \beta_0$ and $\left(\mathrm{K}_1',\mathrm{K}_2\right)$ is in Region III .

3.5.1 General Properties of $\ddot{K}_2(K_1)$ (Region III and Region IV)

We will first show that,

- when K_1 is feasible with $\beta_0 = 0$ then $\ddot{K}_2(K_1) = K_1$,
- when K_1 is feasible with $0 < \beta_0 < 1$ then $\ddot{K}_2(K_1) > K_1$,
- when K_1 is feasible with $0 < \beta_0 < 1, \ddot{K}_2(K_1)$ is a strictly monotone increasing function on its domain
- when $\beta_0 = 1$ there are no feasible contour points.

When $\beta_0 = 0$ then $\ddot{K}_2(K_1) = K_1$

In **Region III** for $\beta_0 = 0$ we require $\Phi\left(K_2 - b\right) - \Phi\left(K_1 - b\right) = 0$. This is satisfied if and only if $K_2 = K_1$. When $\beta_0 = 0$ the contour in Region III is the line segment given by $\frac{a+b}{2} \le K_1 \le b$ and $K_2 = K_1$. The converse is straightforward.

In **Region IV** for $\beta_0 = 0$, we require $0 = -1 + 2\Phi\left(\frac{K_2 - K_1}{2}\right)$. This is satisfied if and only if $K_2 = K_1$. When $\beta_0 = 0$ the contour in Region IV is the line segment given by $K_1 \ge b$ and $K_2 = K_1$.

When $0 < \beta_0 < 1$ then $\ddot{K}_2(K_1) > K_1$

In **Region III** for $0 < \beta_0 < 1$ we require $\Phi\left(K_2 - b\right) - \Phi\left(K_1 - b\right) = \beta_0$. This is equivalent to $\Phi\left(K_2 - b\right) = \beta_0 + \Phi\left(K_1 - b\right)$. Suppose now that $\left(\mathrm{K}_1, \mathrm{K}_2\right)$ is a point satisfying $\Phi\left(K_2 - b\right) = \beta_0 + \Phi\left(K_1 - b\right)$. Since $\beta_0 > 0$, this immediately gives $\Phi\left(K_2 - b\right) > \Phi\left(K_1 - b\right)$ which is equivalent to $K_2 > K_1$. Hence when $0 < \beta_0 < 1$ we have that any solution $(\mathrm{K}_1, \mathrm{K}_2)$ of the contour equation in Region III will immediately have $K_2 > K_1$.

In **Region IV** for $0 < \beta_0 < 1$ we require $\beta_0 = -1 + 2\,\Phi\left(\frac{K_2 - K_1}{2}\right)$. This is equivalent to $K_2 = K_1 + 2\Phi^{-1}\left(\frac{1+\beta_0}{2}\right)$. Now since $\beta_0 > 0$ we have $\Phi^{-1}\left(\frac{1+\beta_0}{2}\right) > 0$ and hence $K_2 > K_1$ in Region IV.

$\ddot{K}_2(K_1)$ **is a strictly monotone increasing function on its domain**

We first consider Region III. Suppose that K_1' ,K_1'' are feasible points on the contour in Region III[15] such that $K_1' < K_1''$ and hence $\Phi(K_1' - b) < \Phi(K_1'' - b)$. Since both points are feasible, we have,

$$\Phi\left(\ddot{K}_2(K_1') - b\right) - \Phi\left(K_1' - b\right) = \beta_0$$

and

$$\Phi\left(\ddot{K}_2(K_1'') - b\right) - \Phi\left(K_1'' - b\right) = \beta_0$$

Combining these gives

$$\Phi\left(\ddot{K}_2(K_1'') - b\right) = \Phi\left(\ddot{K}_2(K_1') - b\right) + \Phi\left(K_1'' - b\right) - \Phi\left(K_1' - b\right)$$

---

15 Saying that K_1 is a feasible value for Region III, means that there exists K_2 such that $(\mathrm{K}_1, \mathrm{K}_2) \in$ Region III and $\beta^*(\mathrm{K}_1, \mathrm{K}_2) = \beta_0$. An analogous definition is used for Region IV.

Since $\Phi\left(K_1' - b\right) < \Phi\left(K_1'' - b\right)$ this gives

$$\Phi\left(\ddot{K}_2(K_1'') - b\right) > \Phi\left(\ddot{K}_2(K_1') - b\right)$$

and hence $\ddot{K}_2(K_1'') > \ddot{K}_2(K_1')$ i.e. $\ddot{K}_2(K_1)$ is strictly monotone increasing. QED

We now consider Region IV. Suppose that K_1', K_1'' are feasible points in Region IV such that $K_1' < K_1''$. Since both points are feasible, we have,

$$\Phi\left(\frac{\ddot{K}_2(K_1'') - K_1''}{2}\right) = \Phi\left(\frac{\ddot{K}_2(K_1') - K_1'}{2}\right)$$

Hence $\ddot{K}_2(K_1'') - K_1'' = \ddot{K}_2(K_1') - K_1'$

This gives $\ddot{K}_2(K_1'') = \ddot{K}_2(K_1') + K_1'' - K_1'$

Since $K_1' < K_1''$, we have $\ddot{K}_2(K_1'') > \ddot{K}_2(K_1')$.

Now consider Region III and Region IV combined. Suppose we have K_1', K_1'' such that $K_1' < K_1''$ where K_1' is feasible for Region III and K_1'' is feasible for Region IV. The contour segments in Region III and Region IV meet the boundary $K_1 + K_2 = 2b$ at $K_1 = b - \Phi^{-1}\left(\frac{\beta_0 + 1}{2}\right)$.

Because $\ddot{K}_2(K_1)$ is strictly monotone increasing in Region III, the set of feasible values of K_1 for contour points in Region III has $K_1 = b - \Phi^{-1}\left(\frac{\beta_0 + 1}{2}\right)$ as its upper boundary and because $\ddot{K}_2(K_1)$ is strictly monotone increasing in Region IV, the set of feasible values of K_1 for contour points in Region IV has $K_1 = b - \Phi^{-1}\left(\frac{\beta_0+1}{2}\right)$ as its lower boundary. Hence we have proved that the set of feasible values of K_1 for contour points in Region III and the set of feasible values of K_1 for contour points in Region IV, intersect only on the boundary $K_1 + K_2 = 2b$. Hence if we have K_1' feasible for Region III and K_1'' feasible for Region IV, we have,

$$K_1' \le b - \Phi^{-1}\left(\frac{\beta_0+1}{2}\right) \text{ and } K_1'' \ge b - \Phi^{-1}\left(\frac{\beta_0+1}{2}\right)$$

Since $K_1' < K_1''$ they cannot both be equal to $b - \Phi^{-1}\left(\frac{\beta_0+1}{2}\right)$ and since $\ddot{K}_2(K_1)$ is strictly monotone increasing in both regions we must therefore have $\ddot{K}_2(K_1'') > \ddot{K}_2(K_1')$. Hence $\ddot{K}_2(K_1)$ is strictly monotone increasing across the region boundary $K_1 + K_2 = 2b$.

Hence we have shown that $\ddot{K}_2(K_1)$ is strictly monotone increasing on its whole domain. QED.

When $\beta_0 = 1$ there are no feasible contour points

In **Region III** for $\beta_0 = 1$ we require $\Phi(K_2 - b) - \Phi(K_1 - b) = 1$. There are no finite values of K_2 and K_1 satisfying this equation. Hence for $\beta_0 = 1$, there are no contour points in Region III.

In **Region IV** for $\beta_0 = 1$ and $\beta_0 = -1+2\Phi\left(\frac{K_2 - K_1}{2}\right)$ which becomes $\Phi\left(\frac{K_2 - K_1}{2}\right) = 1$. Again there are no finite values of K_2 and K_1 satisfying this equation.

In Region III the Contour crosses all lines going to Infinity

Lines going to infinity in Region III have equations of the form $K_1 + K_2 = a+b+\delta$ where $0 \le \delta \le b-a$. The line segment of interest in Region III starts from it rightmost point $\left(\frac{a+b+\delta}{2}, \frac{a+b+\delta}{2}\right)$ and has $-\infty < K_1 \le \frac{a+b+\delta}{2}$. Note that every interior point in Region III lies on such a line for some $0 < \delta < b-a$.

In Region III, $\beta^*(\mathrm{K}_1, \mathrm{K}_2) = \Phi(K_2 - b) - \Phi(K_1 - b)$.

Along the line $K_1 + K_2 = a+b+\delta$ this gives the value

$$\beta^*(\mathrm{K}_1, a+b+\delta-\mathrm{K}_1) = \Phi(a+\delta-K_1) - \Phi(K_1 - b)$$

As $K_1 \to -\infty$, $\beta^*(\mathrm{K}_1, a+b+\delta-\mathrm{K}_1)$ is a continuous monotone increasing function whose limit value =1. In addition when $K_1 = \frac{a+b+\delta}{2}$, we have $\beta^*(\mathrm{K}_1, a+b+\delta-\mathrm{K}_1) = 0$. In other words, along every line $K_1 + K_2 = a+b+\delta$, $\beta^*(\mathrm{K}_1, \mathrm{K}_2)$ takes every value in $[0,1)$. This implies that for all $\beta_0 \in [0,1)$ the contour $\beta^*(\mathrm{K}_1, \mathrm{K}_2) = \beta_0$ has a unique intersection with every line $K_1 + K_2 = a+b+\delta$.

Hence we have that, $\forall \beta_0 : \beta_0 \in [0, 1)$ and $\forall \delta : 0 < \delta < b-a$,

$\exists\ \breve{K}_1\left(\beta_0,\delta\right) \le \frac{a+b+\delta}{2}$ which is unique and if we define $\breve{K}_2 = a+b+\delta-\breve{K}_1\left(\beta_0,\delta\right)$, we have $\beta^*\left(\breve{K}_1\left(\beta_0,\delta\right),\breve{K}_2\right) = \beta_0$.

Note also that when $\breve{K}_1\left(\beta_0,\delta\right) < \frac{a+b+\delta}{2}$ we have $\breve{K}_2 > \frac{a+b+\delta}{2}$. In other words all contour points which are not on $K_2 = K_1$ satisfy $K_2 > K_1$. This was already established earlier without demonstrating the existence of such points.

3.5.2 Intersection of the β^* Contour with the Region III Boundaries.

We now consider the intersection of the β_0 contour with the region boundaries $K_1 + K_2 = a+b$ and $K_1 + K_2 = 2b$ and identify the feasible values of K_1 for the β_0 contour in Regions III and IV. We consider the case $0<\beta_0<1$.

The intersection of $\beta^*\left(\mathrm{K}_1,\mathrm{K}_2\right) = \beta_0$ with $K_1 + K_2 = a+b$ is given by solving $\Phi\left(a-K_1\right)-\Phi\left(K_1-b\right) = \beta_0$. The solution of this equation will be denoted $K_1'\left(\beta_0\right)$ and K_2' will denote $K_2' = a+b-K_1'$. Note that on the line $K_1 + K_2 = a+b$ we need to have $K_1 \le \frac{a+b}{2}$. Hence I need to show $K_1'\left(\beta_0\right) < \frac{a+b}{2}$.

When $\beta_0<1$ the Intersection Point satisfies $K_1'\left(\beta_0\right) < \frac{a+b}{2}$

Proof; We will assume that $\Phi\left(a-K_1\right)-\Phi\left(K_1-b\right) = \beta_0<1$ and that $K_1 \ge \frac{a+b}{2}$ and derive a contradiction. $\Phi\left(a-K_1\right)-\Phi\left(K_1-b\right) = \beta_0$ is equivalent to $\Phi\left(K_1-a\right)+\Phi\left(K_1-b\right)=1-\beta_0<1$.

On the other hand $K_1 \geq \frac{a+b}{2}$ implies,

$$\Phi\left(K_1-a\right) \geq \Phi\left(\frac{a+b}{2}-a\right)=\Phi\left(\frac{b-a}{2}\right) \qquad \text{and}$$

$$\Phi\left(K_1-b\right) \geq \Phi\left(\frac{a+b}{2}-b\right)=\Phi\left(\frac{a-b}{2}\right)=1-\Phi\left(\frac{b-a}{2}\right)$$

These in turn imply,

$$\Phi\left(K_1-a\right)+\Phi\left(K_1-b\right) \geq \Phi\left(\frac{b-a}{2}\right)+1-\Phi\left(\frac{b-a}{2}\right)$$

i.e $\Phi\left(K_1-a\right)+\Phi\left(K_1-b\right) \geq 1$ which is a contradiction of $\Phi\left(K_1-a\right)+\Phi\left(K_1-b\right)=1-\beta_0<1$. Hence for $\beta_0<1$ we have $K_1'\left(\beta_0\right) < \frac{a+b}{2}$. **QED.**

For $\beta_0=0$ and $K_1+K_2=a+b$, the contour equation $\Phi\left(K_2-b\right)-\Phi\left(K_1-b\right)=\beta_0$ requires $\Phi\left(a-K_1\right)-\Phi\left(K_1-b\right)=0$. The solution of this is $K_1=\frac{a+b}{2}$ and hence $K_2=\frac{a+b}{2}$ as well.

We will now show that there is a relationship between K_1' and K_1^*.

The Contour $\beta^*\left(\mathrm{K}_1,\mathrm{K}_2\right)=\beta_0$ meets $K_1+K_2=a+b$ at $K_1^*\left(1-\beta_0\right)$.

The intersection of $\beta^*\left(\mathrm{K}_1,\mathrm{K}_2\right)=\beta_0$ with $K_1+K_2=a+b$ is given by solving $\Phi\left(a-K_1\right)-\Phi\left(K_1-b\right)=\beta_0$, $0\leq\beta_0<1$.

We saw earlier that the intersection of $\alpha^*(\mathrm{K}_1,\mathrm{K}_2) = \alpha_0$ with $K_1 + K_2 = a + b$ is the solution of $\Phi(K_1 - b) + \Phi(K_1 - a) = \alpha_0$, $0 < \alpha_0 \le 1$. It is straightforward to show that when $\alpha_0 = 1 - \beta_0$ these are the same equation .

In other words the contour $\beta^*(\mathrm{K}_1,\mathrm{K}_2) = \beta_0$ and the contour $\alpha^*(\mathrm{K}_1,\mathrm{K}_2) = 1 - \beta_0$ meet $K_1 + K_2 = a + b$ at the same point. This can be said using the notation $K_1^*(\alpha_0)$ defined earlier for the point where $\alpha^*(\mathrm{K}_1,\mathrm{K}_2) = \alpha_0$ meets $K_1 + K_2 = a + b$. In other words we have that the contour $\beta^*(\mathrm{K}_1,\mathrm{K}_2) = \beta_0$ meets $K_1 + K_2 = a + b$ at $K_1^*(1 - \beta_0)$.
In other words, $K_1'(\beta_0) = K_1^*(1 - \beta_0)$.

Intersection of Region III contour with the boundary $K_1 + K_2 = 2b$.

The intersection of $\beta^*(\mathrm{K}_1,\mathrm{K}_2) = \beta_0$ with $K_1 + K_2 = 2b$ is given by solving $\Phi(b - K_1) - \Phi(K_1 - b) = \beta_0$. This is equivalent to $1 - 2\Phi(K_1 - b) = \beta_0$. The solution of this equation will be denoted $K_1^{'b}(\beta_0)$ and $K_2^{'b}$ will denote $K_2^{'b} = 2b - K_1^{'b}$. The solution of this equation is simple and gives $K_1^{'b}(\beta_0) = b - \Phi^{-1}\left(\frac{\beta_0 + 1}{2}\right)$.

Now $\beta_0 \ge 0$ implies $\frac{\beta_0 + 1}{2} \ge \frac{1}{2}$ and this in turn implies $\Phi^{-1}\left(\frac{\beta_0 + 1}{2}\right) \ge 0$.

Hence we have that $K_1^{'b}(\beta_0) \le b$. Note that on the line $K_1 + K_2 = 2b$ we need to have $K_1 \leq b$ to be in the feasible region. In addition, $\beta_0 > 0$ implies $\Phi^{-1}\left(\frac{\beta_0+1}{2}\right) > 0$ and hence $K_1^{'b}\left(\beta_0\right) < b$.

Note also that because $\frac{1+\beta_0}{2}$ and $\frac{1-\beta_0}{2}$ are two probabilities which add to unity, we have that $-\Phi^{-1}\left(\frac{1+\beta_0}{2}\right) = \Phi^{-1}\left(\frac{1-\beta_0}{2}\right)$.

Endpoints of the β_0 Contour in Region III and Feasible Values of K_1

We can summarise the earlier results by saying that, in Region III when $0 < \beta_0 < 1$, the range of feasible K_1 values for the specific β_0 are given by $K_1^*\left(1-\beta_0\right) = K_1'\left(\beta_0\right) \leq K_1 \leq K_1^{'b}\left(\beta_0\right) = b - \Phi^{-1}\left(\frac{\beta_0+1}{2}\right)$ and where $K_1'\left(\beta_0\right)$ is found by solving $\Phi\left(a-K_1\right) - \Phi\left(K_1-b\right) = \beta_0$.

Because of the monotone property is it is clear that for any K_1 satisfying $K_1'\left(\beta_0\right) < K_1 < K_1^{'b}\left(\beta_0\right)$, the contour point $\left(K_1, \ddot{K}_2\left(K_1\right)\right)$ will be an interior point of Region III.

The Feasible Values of K_1 for the β_0 Contour in Region IV

In Region IV the β^* contour $-1+2\,\Phi\left(\frac{K_2-K_1}{2}\right) = \beta_0$ determines K_2 as an implicit function of K_1 for the values K_1 that are possible solutions with the specific value of β_0. Solving the equation for the β_0 contour in Region IV gives the straight line $K_2 = K_1 + 2\Phi^{-1}\left(\frac{\beta_0+1}{2}\right)$. Note that

$\Phi^{-1}\left(\frac{\beta_0+1}{2}\right)>0$ if and only if $\beta_0>0$ and $\Phi^{-1}\left(\frac{\beta_0+1}{2}\right)=0$ if and only if $\beta_0=0$.

We have shown in section 3.1.1 that $\beta^*(K_1, K_2)$ is continuous across the boundary $K_1 + K_2 = 2b$. Here we can verify directly that each contour $\ddot{K}_2(K_1, \beta_0)$ is continuous across this boundary. Solving the contour equation $K_2 = K_1 + 2\Phi^{-1}\left(\frac{1+\beta_0}{2}\right)$ simultaneously with $K_1 + K_2 = 2b$ gives the intersection point

$$K_1 = b - \Phi^{-1}\left(\frac{\beta_0+1}{2}\right) \quad \text{and} \quad K_2 = b + \Phi^{-1}\left(\frac{\beta_0+1}{2}\right).$$

This is the same as the point of intersection of the β_0 contour in Region III and hence $\beta^*(K_1, K_2)$ is continuous across the boundary.

The set of feasible values of K_1 for the β_0 contour in Region IV are $K_1 \geq K_1^{'b}(\beta_0) \equiv b - \Phi^{-1}\left(\frac{\beta_0+1}{2}\right)$.

Summarising we can say that in Region IV the contour starts from the point

$$K_1 = b - \Phi^{-1}\left(\frac{\beta_0+1}{2}\right), \ K_2 = b + \Phi^{-1}\left(\frac{\beta_0+1}{2}\right) \text{ and}$$

is given by $K_2 = K_1 + 2\Phi^{-1}\left(\frac{1+\beta_0}{2}\right)$ for $K_1 \geq b - \Phi^{-1}\left(\frac{\beta_0+1}{2}\right)$.

3.5.3 Properties of the Gradient $\frac{d\ddot{K}_2}{dK_1}$ of the Contour $\ddot{K}_2(K_1,\beta_0)$

In Region III the contour $\Phi(K_2-b)-\Phi(K_1-b)=\beta_0$ determines K_2 as an implicit function of K_1 for the values K_1 that are possible with the specific value of β_0. This function will be denoted $\ddot{K}_2(K_1)$.

We have shown earlier that,

- when K_1 is feasible with $\beta_0 = 0$ then $\ddot{K}_2(K_1)=K_1$,
- when K_1 is feasible with $0<\beta_0<1$ then $\ddot{K}_2(K_1)>K_1$,
- when K_1 is feasible with $0<\beta_0<1, \ddot{K}_2(K_1)$ is a strictly monotone increasing function on its domain
- when $\beta_0 = 1$ there are no feasible contour points.

The β_0 contour in Region III requires the use of the implicit function theorem to derive its properties. From the implicit function theorem, the derivative $\frac{d\ddot{K}_2}{dK_1}$ of $\ddot{K}_2(K_1)$ in $a+b\le K_1+\ddot{K}_2(K_1)\le 2b$ and $\ddot{K}_2(K_1)\ge K_1$ has the equation

$$\frac{d\ddot{K}_2}{dK_1} = \frac{\Phi'(K_1-b)}{\Phi'(\ddot{K}_2-b)} = e^{\frac{1}{2}(K_1+\ddot{K}_2-2b)(\ddot{K}_2-K_1)} >0 \;;\forall K_1, \ddot{K}_2$$

Properties of Terms in the Expression for $\frac{d\ddot{K}_2}{dK_1}$

When $0<\beta_0<1$, the fact that $\ddot{K}_2(K_1)>K_1$ implies that the exponent term $\ddot{K}_2-K_1>0$.

For the β_0 contour points in Region III, we have that $K_1 = b - \Phi^{-1}\left(\frac{\beta_0 + 1}{2}\right)$ if and only if $K_1 + K_2 = 2b$.[16] Since $\ddot{K}_2(K_1)$ is single valued and strictly monotone we have $K_1 < b - \Phi^{-1}\left(\frac{\beta_0 + 1}{2}\right)$ if and only if $K_1 + K_2 < 2b$.

From these considerations we can see that, when $a < b$,

$$\frac{d\ddot{K}_2}{dK_1} = e^{\frac{1}{2}\left(K_1 + \ddot{K}_2 - 2b\right)\left(\ddot{K}_2 - K_1\right)} < 1$$

$$\text{when } a + b \leq K_1 + \ddot{K}_2\left(K_1\right) < 2b$$

We also have that,

$$\frac{d\ddot{K}_2}{dK_1} = e^{\frac{1}{2}\left(K_1 + \ddot{K}_2 - 2b\right)\left(\ddot{K}_2 - K_1\right)} = 1 \quad \text{when } K_1 + \ddot{K}_2\left(K_1\right) = 2b.$$

In other words the contour in Region III has $\frac{d\ddot{K}_2}{dK_1} = 1$ at the moment when it intersects the line $K_1 + K_2 = 2b$.

---

16 $K_1 = b - \Phi^{-1}\left(\frac{\beta_0 + 1}{2}\right)$ and $\Phi\left(K_2 - b\right) - \Phi\left(K_1 - b\right) = \beta_0$

imply that $K_2 = b + \Phi^{-1}\left(\frac{\beta_0 + 1}{2}\right)$.

This means that $\frac{d\ddot{K}_2}{dK_1}$ is continuous across the boundary between Regions III and IV.

Note that the above statements are true for every value of β_0 $0<\beta_0<1$.

The Second Derivative $\frac{d^2\ddot{K}_2}{dK_1^2}$ in Region III

Again from the implicit function theorem the second derivative is given by

$$\frac{d^2\ddot{K}_2}{dK_1^2} = \frac{\Phi'\left(K_1-b\right)}{\left[\Phi'\left(\ddot{K}_2-b\right)\right]^2}\left[\left(\ddot{K}_2-b\right)\Phi'\left(K_1-b\right)-\left(K_1-b\right)\Phi'\left(\ddot{K}_2-b\right)\right]$$

Since $u\,e^{\frac{1}{2}u^2}$ is a monotone increasing function, we have that $\frac{d^2\ddot{K}_2}{dK_1^2}>0$ provided $\ddot{K}_2(K_1)>K_1$. Hence when $\ddot{K}_2(K_1)>K_1$, we have that $\frac{d\ddot{K}_2}{dK_1}$ is increasing.

Note that we have shown earlier that when $0<\beta_0<1$ we have for all feasible K_1, that $\ddot{K}_2(K_1)>K_1$.

Value at the Upper Boundary of the Second Derivative in Region III

When $K_1 + \ddot{K}_2(K_1) = 2b$ we have $\dfrac{d^2\ddot{K}_2}{dK_1^2} = 2(b - \mathrm{K}_1)$.

The value of K_1 for the point of intersection of the contour with

$K_1 + K_2 = 2b$, is just $K_1^{'b}(\beta_0) = b - \Phi^{-1}\left(\dfrac{\beta_0 + 1}{2}\right)$. Hence on the boundary,

we have $\dfrac{d^2\ddot{K}_2}{dK_1^2} = 2\,\Phi^{-1}\left(\dfrac{\beta_0 + 1}{2}\right)$ which is of course > 0 when $\beta_0 > 0$.

The Gradient of the Contour $\ddot{K}_2(K_1)$ in Region IV

The Region IV contour equation is the line $K_2 = K_1 + 2\,\Phi^{-1}\left(\dfrac{\beta_0 + 1}{2}\right)$. Hence

$\dfrac{d\ddot{K}_2}{dK_1} = 1$ throughout Region IV.

We saw earlier that the contour equation for Region III gave

$\dfrac{d^2\ddot{K}_2}{dK_1^2} = 2\,\Phi^{-1}\left(\dfrac{\beta_0 + 1}{2}\right)$ at the intersection with $K_1 + K_2 = 2b$. This

value is of course is > 0 when $\beta_0 > 0$.

In Region IV however $\dfrac{d^2\ddot{K}_2}{dK_1^2} = 0$ which means that $\dfrac{d^2\ddot{K}_2}{dK_1^2}$ is not

continuous across the boundary between Region III and Region IV.

Continuity across the Region Boundary

We saw earlier that the region Region III contour intersects the line $K_1 + K_2 = 2b$ at the same point. Hence we saw that the Region III contour met the Region IV contour on the line $K_1 + K_2 = 2b$.

In addition we saw earlier that the formula for the Region III gradient $\frac{d\ddot{K}_2}{dK_1}$ gives $\frac{d\ddot{K}_2}{dK_1} = 1$ when $K_1 + \ddot{K}_2\left(K_1\right) = 2b$. In other words the contour of Region III has $\frac{d\ddot{K}_2}{dK_1} = 1$ at the moment when it intersects the line $K_1 + K_2 = 2b$. This shows that the contours in the two regions when they meet at $K_1 + K_2 = 2b$ have the same gradient. In other words the total contour defined by different expressions in different regions is nevertheless continuous and differentiable.

In what follows we will use the notation $\ddot{K}_2\left(K_1\right)$, $\frac{d\ddot{K}_2}{dK_1}$, $\frac{d^2\ddot{K}_2}{dK_1^2}$ to speak about the contour in both the Regions III and IV. Remembering of course that in Region IV, $\ddot{K}_2\left(K_1\right)$ is a straight line having $\frac{d\ddot{K}_2}{dK_1} = 1$ and hence $\frac{d^2\ddot{K}_2}{dK_1^2} = 0$ in Region IV. Note however that the second derivative $\frac{d^2\ddot{K}_2}{dK_1^2}$ is not continuous across the boundary since we saw earlier that the Region III expression gives a value $\frac{d^2\ddot{K}_2}{dK_1^2} = 2\,\Phi^{-1}\left(\frac{\beta_0 + 1}{2}\right)$ at the boundary and this is of course > 0 when $\beta_0 > 0$.

The Graph of $\ddot{K}_2\left(K_1\right)$

The above results provide a picture of $\ddot{K}_2\left(K_1\right)$ defined in Region III as the solutions of $\Phi\left(K_2-b\right)-\Phi\left(K_1-b\right)=\beta_0$ for K_1 values satisfying

$$K_1^*\left(1-\beta_0\right) \leq K_1 \leq b-\Phi^{-1}\left(\frac{\beta_0+1}{2}\right)$$

where $K_1^*\left(1-\beta_0\right)$ is found by solving $\Phi\left(a-K_1\right)-\Phi\left(K_1-b\right)=\beta_0$ and $\ddot{K}_2\left(K_1\right)$ is defined in Region IV as $K_2=K_1+2\,\Phi^{-1}\left(\frac{\beta_0+1}{2}\right)$ for $K_1 \geq b-\Phi^{-1}\left(\frac{\beta_0+1}{2}\right)$.

For $0<\beta_0<1$ in Region III, $\ddot{K}_2\left(K_1\right)$ starts from the point $\left(K_1^*\left(1-\beta_0\right),\ddot{K}_2\left(K_1^*\left(1-\beta_0\right)\right)\right)$ with a positive and increasing gradient which is <1 in the subregion $a+b\leq K_1+\ddot{K}_2\left(K_1\right)<2b$ and $=1$ when $K_1+\ddot{K}_2\left(K_1\right)=2b$.

In both Regions III and IV, when $0<\beta_0<1$, the contour remains in the region satisfying $K_2>K_1$ i.e. for $0<\beta_0<1$ there is no intersection with the line $K_2=K_1$.

When $\beta_0=0$, the graph is the half line $K_2=K_1$ having $K_1\geq\frac{a+b}{2}$ (Regions III and IV).

The graph of the contour when $0<\beta_0<1$ is shown in Figure 3 below.

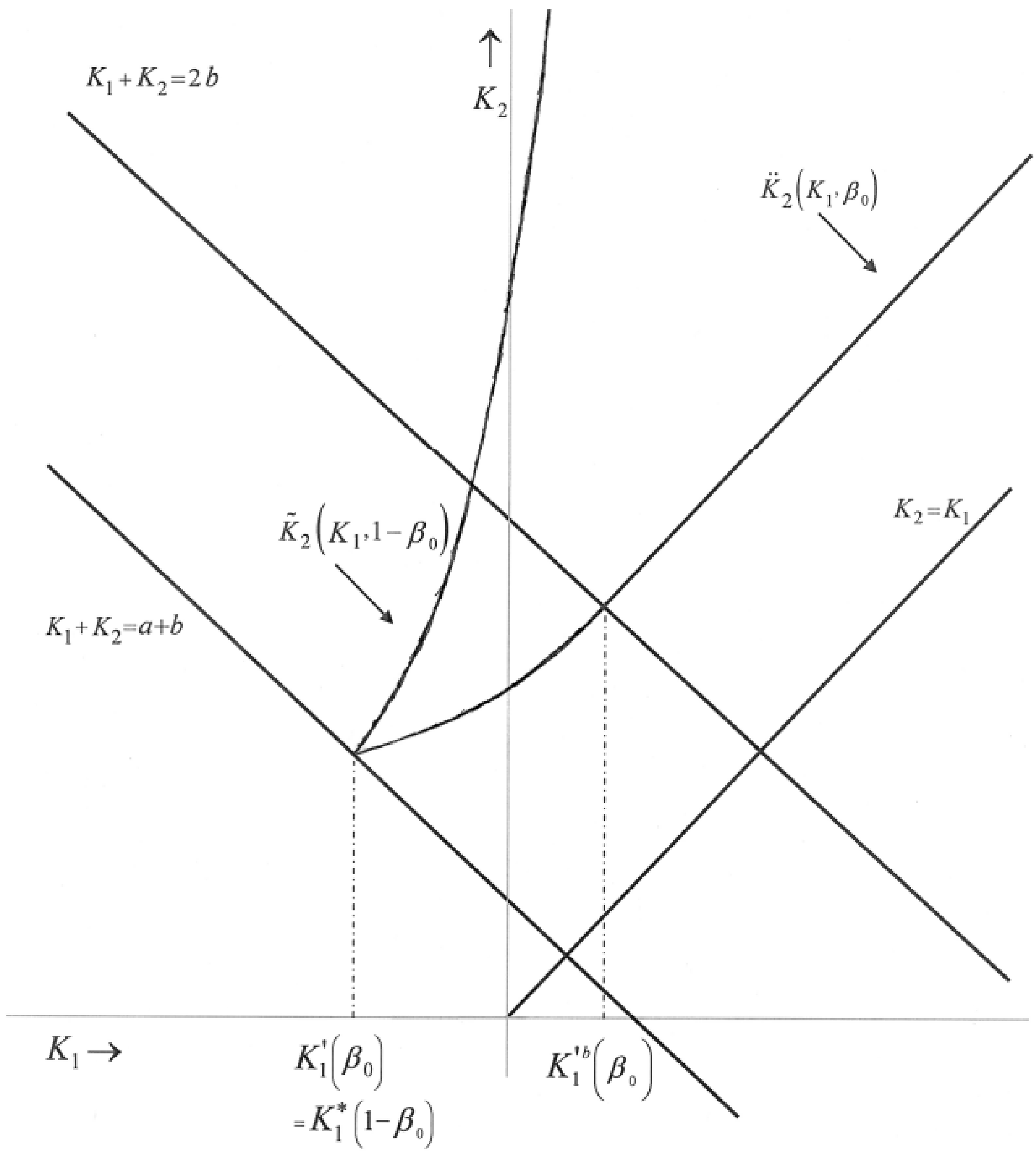


Figure 3: Graph of contour $\ddot{K}_2(K_1, \beta_0)$ also showing $\tilde{K}_2(K_1, 1-\beta_0)$

Comparing the Gradients of $\tilde{K}_2\left(K_1, 1-\beta_0\right)$ and $\ddot{K}_2\left(K_1, \beta_0\right)$.

We saw earlier that the contour $\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right) = \beta_0$ and the contour $\alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right) = 1-\beta_0$ meet $K_1 + K_2 = a + b$ at the same point given by $K_1 = K_1^*\left(1-\beta_0\right)$ (Section 3.5). Here we are using the notation $K_1^*\left(\alpha_0\right)$ defined earlier (2.5.1) for the point where $\alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right) = \alpha_0$ meets $K_1 + K_2 = a + b$.

It is useful to compare the behaviour of the gradients of two such contours which share their point of intersection with $K_1 + K_2 = a + b$. This comparison can only be made of course for the range of K_1 values that are feasible for both contours. The feasible values for $\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right) = \beta_0$ are $K_1 \geq K_1^*\left(1-\beta_0\right)$ and the feasible K_1 values for $\alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right) = 1-\beta_0$ are $K_1^*\left(1-\beta_0\right) \leq K_1 \leq K_1^0\left(1-\beta_0\right)$ where $K_1^0\left(1-\beta_0\right)$ is defined by the equation $\Phi\left(K_1^0 - a\right) = 1-\beta_0$. The two gradients in question are denoted $\tilde{K}_2\left(K_1, 1-\beta_0\right)$ and $\ddot{K}_2\left(K_1, \beta_0\right)$.

Earlier speaking about $\alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right) = \alpha_0$ we have demonstrated that

when $\alpha_0 < 1$ and when $b > a$, $\forall K_1 \in \left[K_1^*\left(\alpha_0\right), K_1^0\left(\alpha_0\right)\right)$,

we have $\dfrac{d\tilde{K}_2}{dK_1} > 1$ (2.5.3.1) and $\dfrac{d^2\tilde{K}_2}{dK_1^2} > 0$ (2.5.3.4) . These statements are true at any point $\left(\mathrm{K}_1, \tilde{K}_2\left(K_1\right)\right)$ of the contour in the region $K_1 + \tilde{K}_2 \geq a + b$. These statements are true for all $\alpha_0 < 1$. Hence they are true when $\alpha_0 = 1-\beta_0 \quad \forall \beta_0 \; 0 < \beta_0 < 1$.

Earlier speaking about $\beta^*(K_1, K_2) = \beta_0$ we have demonstrated that for all $0 < \beta_0 < 1$ and when $b > a$, we have $\frac{d\ddot{K}_2}{dK_1} < 1$ and $\frac{d^2\ddot{K}_2}{dK_1^2} > 0$ in the region $a + b \le K_1 + \ddot{K}_2(K_1) < 2b$ **and we have** $\frac{d\ddot{K}_2}{dK_1} = 1$ and $\frac{d^2\ddot{K}_2}{dK_1^2} = 0$ in the region $K_1 + \ddot{K}_2(K_1) \ge 2b$. Again these statements are true for all $0 < \beta_0 < 1$.

From these we can conclude that for any $\alpha_0 < 1$ and any $0 < \beta_0 < 1$, we have $\frac{d\ddot{K}_2}{dK_1} < \frac{d\tilde{K}_2}{dK_1}$ for all the K_1 values where $\frac{d\tilde{K}_2}{dK_1}$ is defined i.e. for all K_1 satisfying $K_1^*(\alpha_0) \le K_1 < K_1^0(\alpha_0)$. In particular we have this result for $\alpha_0 = 1 - \beta_0$ i.e. with K_1 satisfying $K_1^*(1-\beta_0) \le K_1 < K_1^0(1-\beta_0)$.

Since the two contours $\beta^*(K_1, K_2) = \beta_0$ and $\alpha^*(K_1, K_2) = 1 - \beta_0$ start with equal values at $K_1^*(1-\beta_0)$, we have $\tilde{K}_2(K_1, 1-\beta_0) > \ddot{K}_2(K_1, \beta_0)$ for all K_1 satisfying $K_1^*(1-\beta_0) \le K_1 < K_1^0(1-\beta_0)$.

When $\alpha_0 = 1$, we have $\tilde{K}_2(K_1) = K_1$ (Section 2.5) and when $\beta_0 = 0$, we have $\ddot{K}_2(K_1) = K_1$ (Section 3.3). In other words, the contours $\alpha^*(K_1, K_2) = 1$ and $\beta^*(K_1, K_2) = 0$ both coincide with the line segment $K_2 = K_1$ and $K_1 \ge \frac{a+b}{2}$.

4. Minimising the Function $\sup\left\{ \alpha_d^* , Q\beta_d^* \right\}$

We now wish to find the decision rule **d**: IR → {**AH₀ , RH₀**} that will minimise the function $\sup\left\{ \alpha_d^* , Q\beta_d^* \right\}$. As before, we consider **d**(x) having an acceptance region of the form $\left[K_1\sigma , K_2\sigma\right]$ where $K_1 \le K_2$. In this notation we are interested in finding $\left(K_1, K_2\right)$ to minimise,

$$\sup\left\{ \alpha^*\left(K_1,K_2\right), Q\beta^*\left(K_1,K_2\right) \right\} \text{ where } Q > 0.$$

Symmetry Property

Both $\alpha^*\left(K_1,K_2\right)$ and $\beta^*\left(K_1,K_2\right)$ are symmetric about $K_1 + K_2 = a + b$ and hence $\sup\left\{ \alpha^*\left(K_1,K_2\right), Q\beta^*\left(K_1,K_2\right) \right\}$ is symmetric about $K_1 + K_2 = a + b$. Hence the search for a minimum can be confined to the region $K_2 \ge K_1$ and $K_1 + K_2 \ge a + b$.

Definition of Three subregions

The function $\sup\left\{ \alpha^*\left(K_1,K_2\right), Q\beta^*\left(K_1,K_2\right) \right\}$ divides the region $K_2 \ge K_1$ and $K_1 + K_2 \ge a + b$ into three non-overlapping subregions. These are given by,

$$\left\{\left(K_1,K_2\right): \alpha^*\left(K_1,K_2\right) = Q\beta^*\left(K_1,K_2\right) \right\}$$

$$\left\{\left(K_1,K_2\right): \alpha^*\left(K_1,K_2\right) > Q\beta^*\left(K_1,K_2\right) \right\}$$

$$\left\{\left(K_1,K_2\right): \alpha^*\left(K_1,K_2\right) < Q\beta^*\left(K_1,K_2\right) \right\}$$

To study the function $\sup\left\{ \alpha^*(\mathrm{K}_1,\mathrm{K}_2), \mathrm{Q}\beta^*(\mathrm{K}_1,\mathrm{K}_2) \right\}$ we will identify the location and form of the set ,
$\left\{(\mathrm{K}_1,\mathrm{K}_2): \alpha^*(\mathrm{K}_1,\mathrm{K}_2) = \mathrm{Q}\beta^*(\mathrm{K}_1,\mathrm{K}_2) \right\}$. This set will be referred to as the “gorge” corresponding to Q.

From the implicit function theorem, $\alpha^*(\mathrm{K}_1,\mathrm{K}_2) = \mathrm{Q}\beta^*(\mathrm{K}_1,\mathrm{K}_2)$ defines K_2 as an implicit function of K_1 for the values K_1 that are possible with the specific value of Q. This function will be denoted $\hat{K}_2(K_1)$ for a set of K_1 values that has still to be identified. The values of the gorge function $\hat{K}_2(K_1)$ also depend on the value of Q. When it is desirable to emphasise this fact, the notation $\hat{K}_2(K_1,\mathrm{Q})$ will be used.

Since we are optimizing in the region $K_2 \geq K_1$ and $K_1 + K_2 \geq a + b$, the gorge points **of interest** will be those having $K_1 + \hat{K}_2(K_1) \geq a + b$ and $\hat{K}_2 \geq K_1$.

4.1 Location of the Gorge Points $\alpha^*(\mathrm{K}_1,\mathrm{K}_2) = \mathrm{Q}\beta^*(\mathrm{K}_1,\mathrm{K}_2)$

We can write the gorge equation in terms of the earlier expressions for $\alpha^*(\mathrm{K}_1,\mathrm{K}_2)$ and $\beta^*(\mathrm{K}_1,\mathrm{K}_2)$.

These earlier equations were,

$$\alpha^*(\mathrm{K}_1,\mathrm{K}_2) = 1 + \Phi(K_1 - a) - \Phi(K_2 - a) \qquad \text{(from Section 2.2)}$$

and for $\beta^*(\mathrm{K}_1,\mathrm{K}_2)$ from Section 3.1.1 we have,

In Region III: $a + b \leq K_1 + K_2 \leq 2b$;

$$\beta^{*}(\mathrm{K}_1, \mathrm{K}_2) \quad = \Phi(K_2 - b) - \Phi(K_1 - b)$$

and In Region IV: $2b \leq K_1 + K_2$;

$$\beta^{*}(\mathrm{K}_1, \mathrm{K}_2) = -1 + 2\,\Phi\left(\frac{K_2 - K_1}{2}\right)$$

Hence the gorge equation becomes,

In Region III: $a + b \leq K_1 + K_2 \leq 2b$;

$$1 + \Phi(K_1 - a) - \Phi(K_2 - a) = \mathrm{Q}\Phi(K_2 - b) - \mathrm{Q}\Phi(K_1 - b)$$

In Region IV: $2b \leq K_1 + K_2$;

$$1 + \Phi(K_1 - a) - \Phi(K_2 - a) = -\mathrm{Q} + 2\mathrm{Q}\,\Phi\left(\frac{K_2 - K_1}{2}\right)$$

4.1.1 The Gorge Points satisfy $\hat{K}_2(K_1) > \mathrm{K}_1$

In Region III i.e. $a + b \leq K_1 + K_2 \leq 2b$, **the gorge equation is**,

$$1 + \Phi(K_1 - a) - \Phi(K_2 - a) = \mathrm{Q}\Phi(K_2 - b) - \mathrm{Q}\Phi(K_1 - b)$$

Suppose first that $K_2 < K_1$ and hence the Left Hand Side is >1. But if $K_2 < K_1$ the Right Hand Side is <0 (Q is >0). Hence we have a contradiction.

Now suppose that $K_2 = K_1$. The RHS is $=1$ and the LHS is $=0$. Hence we have a contradiction.

Hence we have that $\hat{K}_2(K_1) > \mathrm{K}_1$ for any point $\left(\mathrm{K}_1, \hat{K}_2(K_1)\right)$ satisfying the gorge equation of Region III.

In Region IV i.e. $2b \le K_1 + K_2$, t**he gorge equation** is,

$$1 + \Phi(K_1 - a) - \Phi(K_2 - a) = -\mathrm{Q} + 2\mathrm{Q}\,\Phi\left(\frac{K_2 - K_1}{2}\right)$$

Suppose first that $K_2 < K_1$ and hence the LHSide is >1. But if $K_2 < K_1$ then $\Phi\left(\frac{K_2 - K_1}{2}\right) < \frac{1}{2}$ and hence the RHSide is <0. Hence we have a contradiction.

Now suppose that $K_2 = K_1$. The RHS is $=1$ and the LHS is $=0$. Hence we have a contradiction.

Hence we have that $\hat{K}_2(K_1) > \mathrm{K}_1$ for any point $\left(\mathrm{K}_1, \hat{K}_2(K_1)\right)$ satisfying the gorge equation of Region IV. QED.

Note that this is true $\forall \mathrm{Q}:\ \mathrm{Q} > 0$ i.e. if we define $\alpha_0^{**}(\mathrm{Q}) = \frac{\mathrm{Q}}{1+\mathrm{Q}} < 1$, it is true $\forall \alpha_0^{**}(\mathrm{Q})$.

4.1.2 The intersection of the gorge with $K_1 + K_2 = a + b$.

Part I: We must first examine whether $\alpha^*(\mathrm{K}_1,\mathrm{K}_2) = \mathrm{Q}\beta^*(\mathrm{K}_1,\mathrm{K}_2)$ has any intersection with $K_1 + K_2 = a + b$. **If it does** then such a K_1 must satisfy

$$\alpha^*\left(\mathrm{K}_1, a + b - K_1\right) = \mathrm{Q}\beta^*\left(\mathrm{K}_1, a + b - K_1\right).$$

We can substitute in this equation the formulae derived earlier for $\alpha^*\left(\mathrm{K}_1, a + b - K_1\right)$ and $\beta^*\left(\mathrm{K}_1, a + b - K_1\right)$. These earlier formulae are,

$$\alpha^*\left(K_1, a+b-K_1\right) = 1+\Phi\left(K_1-a\right)-\Phi\left(b-K_1\right)$$

and

$$\beta^*\left(K_1, a+b-K_1\right) \quad = \Phi\left(a-K_1\right)-\Phi\left(K_1-b\right)$$

Substituting these gives the equation that K_1 must satisfy,

$$Q\Phi\left(a-K_1\right)-Q\Phi\left(K_1-b\right) = 1+\Phi\left(K_1-a\right)-\Phi\left(b-K_1\right)$$

This can be rewritten,

$$\Phi\left(K_1-a\right)+\Phi\left(K_1-b\right) = \frac{Q}{1+Q} \quad .$$

We immediately recognise this equation as the defining equation of $K_1^*\left(\alpha_0\right)$. Earlier in Section 2.5.1, $K_1^*\left(\alpha_0\right)$ was defined as the solution of $\alpha^*\left(K_1^*, a+b-K_1^*\right) = \alpha_0$ which produced the same equation. The only difference is that now α_0 is replaced by $\frac{Q}{1+Q}$.

In other words we have established that the Q gorge intersects the boundary $K_1+K_2=a+b$ at the point given by

$$K_1 = K_1^*\left(\tfrac{Q}{1+Q}\right) \text{ and } K_2 = a+b-K_1^*\left(\tfrac{Q}{1+Q}\right)$$

Note that $Q>0$ implies $\frac{Q}{1+Q}<1$ which in turn implies $K_1^*\left(\frac{Q}{1+Q}\right)<\frac{a+b}{2}$ (see Section 2.5.1).

If we let $\alpha_0^{**}(\mathrm{Q})$ denote the value of $\alpha^*(\mathrm{K}_1,\mathrm{K}_2)$ at the point where the gorge $\alpha^*(\mathrm{K}_1,\mathrm{K}_2) = \mathrm{Q}\beta^*(\mathrm{K}_1,\mathrm{K}_2)$ meets $K_1 + K_2 = a + b$, we have $\alpha_0^{**}(\mathrm{Q}) = \frac{\mathrm{Q}}{1+\mathrm{Q}} < 1$.

Alternatively if we use α_0^{**} as the defining parameter of the gorge, we have $\mathrm{Q} = \frac{\alpha^{**}}{1-\alpha^{**}}$. This inverse relationship however can only be defined for $\alpha_0^{**}(\mathrm{Q}) < 1$. This is not a problem since the concept of gorge is only defined for $\forall\mathrm{Q}:\ \mathrm{Q} > 0$ and hence is only defined for $\forall\mathrm{Q}:\ \alpha_0^{**}(\mathrm{Q}) < 1$.

From the gorge equation we can also write

$$\beta^*\left(K_1^*\left(\frac{\mathrm{Q}}{1+\mathrm{Q}}\right), a+b - K_1^*\left(\frac{\mathrm{Q}}{1+\mathrm{Q}}\right)\right) = \frac{1}{1+\mathrm{Q}}$$

Note that the same results can be proved by another argument.

Earlier in Section 3.1.3, we have shown that at any point along the boundary $K_1 + K_2 = a + b$, we have

$$\beta^*(\mathrm{K}_1, a+b - K_1) = 1 - \alpha^*(\mathrm{K}_1, a+b - K_1),\ \forall\ \mathrm{K}_1\ .$$

At the point where the gorge $\alpha^*(\mathrm{K}_1,\mathrm{K}_2) = \mathrm{Q}\beta^*(\mathrm{K}_1,\mathrm{K}_2)$ intersects the boundary $K_1 + K_2 = a + b$, we must have,

$$\alpha^*(K_1, a+b-K_1) = \mathrm{Q}\beta^*(K_1, a+b-K_1)$$

For an intersection point, these two equations must be simultaneously true. Solving simultaneously gives,

$$\beta^{*}\left(\mathrm{K}_1, a+b-K_1\right) = 1-\mathrm{Q}\,\beta^{*}\left(K_1, a+b-K_1\right) .$$

This gives

$$\beta^{*}\left(\mathrm{K}_1, a+b-K_1\right) = \frac{1}{1+\mathrm{Q}} \quad \text{for the “Q intersection point”.}$$

and

$$\alpha^{*}\left(\mathrm{K}_1, a+b-K_1\right) = \frac{\mathrm{Q}}{1+\mathrm{Q}} \quad \text{at the Q intersection point.}$$

Hence we have as before that $K_1^{*}\left(\frac{\mathrm{Q}}{1+\mathrm{Q}}\right)$ satisfies,

$$\beta^{*}\left(K_1^{*}\left(\frac{\mathrm{Q}}{1+\mathrm{Q}}\right), a+b-K_1^{*}\left(\frac{\mathrm{Q}}{1+\mathrm{Q}}\right)\right) = \frac{1}{1+\mathrm{Q}} .$$

and

$$\alpha^{*}\left(K_1^{*}\left(\frac{\mathrm{Q}}{1+\mathrm{Q}}\right), a+b-K_1^{*}\left(\frac{\mathrm{Q}}{1+\mathrm{Q}}\right)\right) = \frac{\mathrm{Q}}{1+\mathrm{Q}} .$$

$\alpha_0^{**}(\mathrm{Q})$ as a Function of the “Penalty” Costs p_1 and p_2

Remember also that in starting with the loss function formulation, we defined $\mathrm{Q} = \frac{\mathrm{p}_2}{\mathrm{p}_1}$. Substituting this in the expression for $\alpha_0^{**}(\mathrm{Q})$ we can write

$$\alpha_0^{**}(\mathrm{p}_1, \mathrm{p}_2) = \frac{\mathrm{p}_2}{\mathrm{p}_1 + \mathrm{p}_2}.$$

4.1.3 The Point of Intersection of the gorge with $K_1 + K_2 = 2b$

The first fact to notice is that the gorge segments in the separate regions meet the boundary $K_1 + K_2 = 2b$ at the same value of K_1. This can be established by substituting $K_2 = 2b - K_1$ in both equations and seeing that the resulting equation in K_1 is the same in both cases. The substitution gives,

In Region III: $a + b \leq K_1 + K_2 \leq 2b$;

$$1 + \Phi(K_1 - a) - \Phi(2b - K_1 - a) = \mathrm{Q}\Phi(2b - K_1 - b) - \mathrm{Q}\Phi(K_1 - b)$$

which is equivalent to

$$1 + \Phi(K_1 - a) - \Phi(2b - K_1 - a) = \mathrm{Q}\Phi(b - K_1) - \mathrm{Q}\Phi(K_1 - b)$$

which is equivalent to

$$1 + \Phi(K_1 - a) - \Phi(2b - K_1 - a) = \mathrm{Q}\Phi(b - K_1) - \mathrm{Q} + \mathrm{Q}\Phi(b - K_1)$$

or equivalently

$$1 + \Phi(K_1 - a) - \Phi(2b - K_1 - a) = 2\mathrm{Q}\Phi(b - K_1) - \mathrm{Q}$$

In Region IV: $2b \leq K_1 + K_2$;

$$1 + \Phi(K_1 - a) - \Phi(2b - K_1 - a) = -\mathrm{Q} + 2\mathrm{Q}\,\Phi\left(\frac{2b - K_1 - K_1}{2}\right)$$

which is equivalent to

$$1 + \Phi(K_1 - a) - \Phi(2b - K_1 - a) = -\mathrm{Q} + 2\mathrm{Q}\,\Phi(b - K_1)$$

The two equations in K_1 are the same. Hence the gorge in Region III and the gorge in Region IV intersect $K_1 + K_2 = 2b$ at the same point. **QED.**

Notation for Point of Intersection with $K_1 + K_2 = 2b$

Intersecting with the line $K_1 + K_2 = 2b$ the gorge must satisfy

$$1 + \Phi(K_1 - a) - \Phi(\ 2b - K_1 - a) = -\mathrm{Q} + 2\mathrm{Q}\,\Phi(b - K_1)$$

which is equivalent to

$$\Phi(K_1 - a) + \Phi(K_1 - 2b + a) + 2\mathrm{Q}\,\Phi(K_1 - b) - \mathrm{Q} = 0 \quad .$$

The range of feasible values of K_1 along $K_1 + K_2 = 2b$ are $K_1 \leq b$.

At the point (b, b), $\Phi(K_1 - a) + \Phi(K_1 - 2b + a) + 2\mathrm{Q}\,\Phi(K_1 - b) - \mathrm{Q}$ takes the value 1, i.e. > 0.

Moving along the line $K_1 + K_2 = 2b$ with $K_1 \to -\infty$, the value of $\Phi(K_1 - a) + \Phi(K_1 - 2b + a) + 2\mathrm{Q}\,\Phi(K_1 - b) - \mathrm{Q}$ decreases continuously and monotonely to $-\mathrm{Q}$ (the derivative is always positive).

Hence a unique value of K_1 exists with $K_1 < b$ such that,

$$\Phi(K_1 - a) + \Phi(K_1 - 2b + a) + 2\mathrm{Q}\,\Phi(K_1 - b) - \mathrm{Q} = 0$$

In other words, for every $\mathrm{Q} > 0$, the Q gorge meets with the line $K_1 + K_2 = 2b$ at $K_1 = K_1^{**}$ where K_1^{**} is the unique solution of

$$\Phi\left(K_1 - a\right) + \Phi\left(K_1 - 2b + a\right) + 2\mathrm{Q}\,\Phi\left(K_1 - b\right) - \mathrm{Q} = 0 \;.$$

The value K_1^{**} depends on Q and when this is important the notation $K_1^{**}(\mathrm{Q})$ will be used. Note that since $K_1^{**}(\mathrm{Q})$ is a gorge point, we have from section 4.1.1 that $K_1^{**}(\mathrm{Q}) < b \;\; \forall \mathrm{Q}:\; \mathrm{Q} > 0$.

It is a property of $K_1^{**}(\mathrm{Q})$ that $K_1^{**}(\mathrm{Q}) > K_1^{*b}\left(\frac{\mathrm{Q}}{1+\mathrm{Q}}\right)$

Remember from section 2.5.2 that for $\alpha_0 < 1$, $K_1^{*b}(\alpha_0) < b$ is the K_1 value where $\alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right) = \alpha_0$ meets $K_1 + K_2 = 2b$. Hence we have that,

$$K_1^{*b}\left(\frac{\mathrm{Q}}{1+\mathrm{Q}}\right) \quad \text{satisfies } \Phi\left(K_1^{*b} + a - 2b\right) + \Phi\left(K_1^{*b} - a\right) = \frac{\mathrm{Q}}{1+\mathrm{Q}}$$

$$\text{and } K_1^{*b}\left(\frac{\mathrm{Q}}{1+\mathrm{Q}}\right) < b \;\; \forall \mathrm{Q}:\; \mathrm{Q} > 0 .$$

We also have that $K_1^{**}(\mathrm{Q}) < b$ satisfies,

$$\Phi\left(K_1^{**} - a\right) + \Phi\left(K_1^{**} - 2b + a\right) + 2\mathrm{Q}\,\Phi\left(K_1^{**} - b\right) - \mathrm{Q} = 0$$

Equating "Q = Q" the simultaneous equations give,

$$\Phi\left(K_1^{**} - a\right) + \Phi\left(K_1^{**} - 2b + a\right) + 2\mathrm{Q}\,\Phi\left(K_1^{**} - b\right)$$

$$= \left(1+\mathrm{Q}\right)\Phi\left(K_1^{*b} + a - 2b\right) + \left(1+\mathrm{Q}\right)\Phi\left(K_1^{*b} - a\right)$$

This can be written,

$$\Phi\left(K_1^{**}-a\right)+\Phi\left(K_1^{**}-2b+a\right)-\Phi\left(K_1^{*b}+a-2b\right)-\Phi\left(K_1^{*b}-a\right)$$

$$=\mathrm{Q}\left[\Phi\left(K_1^{*b}+a-2b\right)+\Phi\left(K_1^{*b}-a\right)-2\Phi\left(K_1^{**}-b\right)\right]$$

We can now prove that $K_1^{**}(\mathrm{Q})>\ K_1^{*b}\left(\frac{\mathrm{Q}}{1+\mathrm{Q}}\right)\qquad\forall\mathrm{Q}:\ \mathrm{Q}>0$**.**

Assume instead that $K_1^{**}(\mathrm{Q})<K_1^{*b}\left(\frac{\mathrm{Q}}{1+\mathrm{Q}}\right)$ which implies that

$\Phi\left(K_1^{*b}-a\right)>\Phi\left(K_1^{**}-a\right)$ and $\Phi\left(K_1^{*b}-2b+a\right)>\Phi\left(K_1^{**}-2b+a\right)$.

Hence if $K_1^{**}<K_1^{*b}$, the Left Hand Side (LHS) of the equation is <0.

We will now show that under the assumption $K_1^{**}<K_1^{*b}$, the Right Hand Side (RHS) is >0. This is a contradiction of LHS <0.

Derivation of contradiction.

The RHS is $\mathrm{Q}\left[\Phi\left(K_1^{*b}+a-2b\right)+\Phi\left(K_1^{*b}-a\right)-2\Phi\left(K_1^{**}-b\right)\right]$

To derive a contradiction it would be sufficient to show that

$$\Phi\left(K_1^{*b}+a-2b\right)+\Phi\left(K_1^{*b}-a\right)-2\Phi\left(K_1^{**}-b\right)>0$$

This can also be written,

$$\Phi\left(K_1^{*b}-a\right)-\Phi\left(K_1^{**}-b\right)>\Phi\left(K_1^{**}-b\right)-\Phi\left(K_1^{*b}+a-2b\right)$$

To prove this for all $K_1^{**} < K_1^{*b}$ it would be sufficient to show that the infimum value with respect to K_1^{**} of the LHS is $>$ the supremum value with respect to K_1^{**} of the RHS. The infimum value of the LHS is given by $K_1^{**} = K_1^{*b}$ and the supremum value of the RHS is also given by $K_1^{**} = K_1^{*b}$.

Hence it would be sufficient to show

$$\Phi\left(K_1^{*b} - a\right) - \Phi\left(K_1^{*b} - b\right) > \Phi\left(K_1^{*b} - b\right) - \Phi\left(K_1^{*b} + a - 2b\right)$$

Each side of the inequality is a standard Gaussian integral over an interval of length $b-a$ and the upper limit of the RHS interval (i.e. $K_1^{*b} - b$) is the lower limit of the LHS interval. For such Gaussian integrals, it is straightforward to show that,

- LHS > RHS if and only if $K_1^{*b} - b < 0$
- LHS = RHS if and only if $K_1^{*b} - b = 0$ (not feasible value of K_1^{*b})
- LHS < RHS if and only if $K_1^{*b} - b > 0$ (not feasible values of K_1^{*b})

Because $K_1^{*b} - b < 0$, we therefore have that,

$$\Phi\left(K_1^{*b} - a\right) - \Phi\left(K_1^{*b} - b\right) > \Phi\left(K_1^{*b} - b\right) - \Phi\left(K_1^{*b} + a - 2b\right)$$

and hence $\Phi\left(K_1^{*b} + a - 2b\right) + \Phi\left(K_1^{*b} - a\right) - 2\Phi\left(K_1^{*b} - b\right) > 0$.

Since by assumption $\Phi\left(K_1^{*b} - b\right) > \Phi\left(K_1^{**} - b\right)$, we have that

$$\Phi\left(K_1^{*b}+a-2b\right)+\Phi\left(K_1^{*b}-a\right)-2\Phi\left(K_1^{**}-b\right)>0.$$

Hence we have shown that for $\forall \mathrm{Q}:\ \mathrm{Q}\ >\ 0$, the assumption $K_1^{**}<K_1^{*b}$ leads to a contradiction.

The Other Alternative

Now consider the assumption $K_1^{**}=K_1^{*b}$ along with the equation

$$\Phi\left(K_1^{**}-a\right)+\Phi\left(K_1^{**}-2b+a\right)-\Phi\left(K_1^{*b}+a-2b\right)-\Phi\left(K_1^{*b}-a\right)$$

$$=\mathrm{Q}\left[\Phi\left(K_1^{*b}+a-2b\right)+\Phi\left(K_1^{*b}-a\right)-2\Phi\left(K_1^{**}-b\right)\right]$$

When $K_1^{**}=K_1^{*b}$, the LHS = 0. When $K_1^{**}=K_1^{*b}$ the RHS > 0 .

The proof is similar to the preceding case but simpler because we can immediately substitute $K_1^{**}=K_1^{*b}$. Hence for $\forall \mathrm{Q}:\ \mathrm{Q}\ >\ 0$, the assumption $K_1^{**}=K_1^{*b}$ also leads to a contradiction.

Hence for $\forall \mathrm{Q}:\ \mathrm{Q}\ >\ 0$ since both alternatives lead to contradiction we have that $K_1^{**}(\mathrm{Q})>\ K_1^{*b}\left(\frac{\mathrm{Q}}{1+\mathrm{Q}}\right)$ **. QED.**

4.1.4 The Gorge $\hat{K}_2(K_1)$ is a monotone increasing function

In Region III i.e. $a+b \le K_1 + K_2 \le 2b$, **the gorge equation** is,

$$1+\Phi(K_1-a)-\Phi(K_2-a)=\mathrm{Q}\Phi(K_2-b)-\mathrm{Q}\Phi(K_1-b)$$

We have shown earlier (in 4.1.1) that $\hat{K}_2(K_1) > \mathrm{K}_1$ for any point $(\mathrm{K}_1, \hat{K}_2(K_1))$ satisfying the gorge equation of Region III.

Suppose that K_1', K_1'' are feasible gorge points in Region III such that $K_1' < K_1''$ and hence $\Phi(K_1'-a) < \Phi(K_1''-a)$. Since both points are feasible, we have

$$1+\Phi(K_1'-a)-\Phi(\hat{K}_2(K_1')-a)=\mathrm{Q}\Phi(\hat{K}_2(K_1')-b)-\mathrm{Q}\Phi(K'_1-b)$$

and

$$1+\Phi(K_1''-a)-\Phi(\hat{K}_2(K_1'')-a)=\mathrm{Q}\Phi(\hat{K}_2(K_1'')-b)-\mathrm{Q}\Phi(K_1''-b)$$

Subtracting these two equations gives

$$\begin{aligned}\Phi(\hat{K}_2(K_1'')-a)-\Phi(\hat{K}_2(K_1')-a)&\\ = \quad & \Phi(K_1''-a) - \Phi(K_1'-a)\\ & - \mathrm{Q}\left[\Phi(\hat{K}_2(K_1'')-b)-\Phi(\hat{K}_2(K_1')-b)\right]\\ & +\mathrm{Q}\left[\Phi(K_1''-b)-\Phi(K_1'-b)\right]\end{aligned}$$

If we now assume that $\hat{K}_2(K''_1)=\hat{K}_2(K_1')$ the above equation gives a contradiction (i.e. LHS=0 and RHS>0).

Likewise if we assume that $\hat{K}_2(K''_1) < \hat{K}_2(K'_1)$ we get a contradiction (i.e. LHS<0 and RHS>0).

Hence we have that for all K'_1, K''_1 that are coordinates of gorge points in Region III and such that $K'_1 < K''_1$, we have $\hat{K}_2(K''_1) > \hat{K}_2(K'_1)$. QED

In Region IV i.e. $2b \le K_1 + K_2$, t**he gorge equation** is,

$$1 + \Phi(K_1 - a) - \Phi(K_2 - a) = -Q + 2Q\,\Phi\left(\frac{K_2 - K_1}{2}\right)$$

We have shown earlier (in 4.1.1) that $\hat{K}_2(K_1) > \mathrm{K}_1$ for any point $\left(\mathrm{K}_1, \hat{K}_2(K_1)\right)$ satisfying the gorge equation of Region IV.

Suppose that K'_1, K''_1 are feasible points in Region IV such that $K'_1 < K''_1$ and hence $\Phi(K'_1 - a) < \Phi(K''_1 - a)$. Since both points are feasible, we have,

$$1 + \Phi(K'_1 - a) - \Phi(\hat{K}_2(K'_1) - a) = -Q + 2Q\,\Phi\left(\frac{\hat{K}_2(K'_1) - K'_1}{2}\right)$$

and

$$1 + \Phi(K''_1 - a) - \Phi(\hat{K}_2(K''_1) - a) = -Q + 2Q\,\Phi\left(\frac{\hat{K}_2(K''_1) - K''_1}{2}\right)$$

Subtracting these two equations gives

$$\Phi\left(\hat{K}_2(K_1'')-a\right)-\Phi\left(\hat{K}_2(K_1')-a\right)$$
$$= \quad \Phi\left(K_1''-a\right)-\Phi\left(K_1'-a\right)$$
$$+\ 2\mathrm{Q}\left[\Phi\left(\frac{\hat{K}_2\left(K_1'\right)-K_1'}{2}\right)-\Phi\left(\frac{\hat{K}_2\left(K_1''\right)-K_1''}{2}\right)\right]$$

If we now assume that $\hat{K}_2\left(K''_1\right)=\hat{K}_2\left(K'_1\right)$ the above equation produces a contradiction (i.e. LHS=0 and RHS>0).

Likewise if we assume that $\hat{K}_2\left(K''_1\right)<\hat{K}_2\left(K'_1\right)$ we get a contradiction (i.e. LHS<0 and RHS>0).

Hence we have that for all K'_1, K''_1 that are coordinates of gorge points in Region IV and such that $K'_1 < K''_1$, we have $\hat{K}_2\left(K''_1\right)>\hat{K}_2\left(K'_1\right)$. QED

Note also that since the gorge in Region III and the gorge in Region IV meet $K_1+K_2=2b$ at the same point $\left(K_1^{**}(\mathrm{Q}), 2b-K_1^{**}(\mathrm{Q})\right)$, it is straightforward to show that for all K'_1, K''_1 that are coordinates of gorge points in Region III and Region IV respectively and such that $K'_1 < K''_1$, we have $\hat{K}_2\left(K''_1\right)>\hat{K}_2\left(K'_1\right)$. They cannot both be equal to $K_1^{**}(\mathrm{Q})$.

4.1.5 $\alpha^*(K_1, K_2)=\frac{\mathrm{Q}}{1+\mathrm{Q}}$ intersects the gorge only on $K_1+K_2=a+b$

The term “intersection point” is used here to refer to any point $\left(\mathrm{K}_1, \mathrm{K}_2\right)$ which satisfies $\alpha^*(K_1, K_2)=\frac{\mathrm{Q}}{1+\mathrm{Q}}$ and that also satisfies the gorge equation

$\alpha^*(K_1,K_2) = Q\beta^*(K_1,K_2)$. Any such an intersection point must have $K_2 > K_1$ since all points on $\alpha^*(K_1,K_2) = \frac{Q}{1+Q}$ have $K_2 > K_1$ because $\frac{Q}{1+Q} < 1$ (see Section 2.5.).

The implicit function of the gorge is defined by,

In Region III: $a+b \le K_1 + K_2 \le 2b$;

$$1+\Phi(K_1-a)-\Phi(K_2-a) = Q\Phi(K_2-b) - Q\Phi(K_1-b)$$

In Region IV: $2b \le K_1 + K_2$;

$$1+\Phi(K_1-a)-\Phi(K_2-a) = -Q + 2Q\Phi\left(\frac{K_2-K_1}{2}\right)$$

The implicit function for the contour $\alpha^*(K_1,K_2) = \frac{Q}{1+Q}$ is

$$1+\Phi(K_1-a)-\Phi(K_2-a) = \frac{Q}{1+Q}$$

First we characterise intersection points in Region III.

Subtracting $1+\Phi(K_1-a)-\Phi(K_2-a) = \frac{Q}{1+Q}$ from the gorge equation gives $\Phi(K_2-b)-\Phi(K_1-b) = \frac{1}{1+Q}$. The contour $\alpha^*(K_1,K_2) = \frac{Q}{1+Q}$ can also be written $\Phi(K_2-a)-\Phi(K_1-a) = \frac{1}{1+Q}$. Subtracting these two equations gives $\Phi(K_2-a)-\Phi(K_1-a) = \Phi(K_2-b)-\Phi(K_1-b)$. Any intersection point must satisfy this equation. We will see below that there are also solutions that are not intersection points. We will characterize all solutions of the equation and show that intersection points having $K_2 > K_1$ necessarily satisfy $K_1 + K_2 = a + b$.

Note that any (K_1, K_2) with $K_2 > K_1$ that satisfies this equation will be a solution of the simultaneous equations for some Q because the value of $\Phi(K_2 - a) - \Phi(K_1 - a)$ determines the value Q. The set of all (K_1, K_2) satisfying $\Phi(K_2 - a) - \Phi(K_1 - a) = \Phi(K_2 - b) - \Phi(K_1 - b)$ contains the union over Q of all solutions of the equations for each single value $Q: \ 0 < Q < +\infty$. Note however that this equation has $K_2 = K_1$ as a solution but $K_2 = K_1$ is not a solution of the gorge nor of the contour $\alpha^*(K_1, K_2) = \frac{Q}{1+Q}$ for any finite value of Q.

The equation $\Phi(K_2 - a) - \Phi(K_1 - a) = \Phi(K_2 - b) - \Phi(K_1 - b)$ can be written as

$$\frac{1}{\sqrt{2\pi}} \int_{K_1-a}^{K_2-a} e^{-\frac{1}{2}u^2} du \quad = \quad \frac{1}{\sqrt{2\pi}} \int_{K_1-b}^{K_2-b} e^{-\frac{1}{2}u^2} du$$

It is a straightforward property of these integrals (see Appendix A) that the only possible solutions of this in the region $K_2 \geq K_1$ when $a \neq b$, are given by all points (K_1, K_2) satisfying $K_2 = K_1$ and all points (K_1, K_2) satisfying $K_2 > K_1$ and $K_1 - a = -(K_2 - b)$ which is just $K_1 + K_2 = a + b$.

Hence there are no solutions with $K_2 > K_1$ except the solution where both gorge and contour meet on $K_1 + K_2 = a + b$. As shown earlier this point given by

$K_1 = K_1^*\left(\frac{Q}{1+Q}\right)$ and $K_2 = a + b - K_1^*\left(\frac{Q}{1+Q}\right)$.

Now we characterise possible intersection points in Region IV.

Subtracting the contour equation $1+\Phi(K_1-a)-\Phi(K_2-a)=\frac{Q}{1+Q}$ from the gorge equation $1+\Phi(K_1-a)-\Phi(K_2-a)= -Q+2Q\,\Phi\left(\frac{K_2-K_1}{2}\right)$ gives $\Phi\left(\frac{K_2-K_1}{2}\right)=\frac{Q+2}{2Q+2}$. Substituting $K_2=K_1$ in this equation gives a contradiction. Hence there are no solutions having $K_2=K_1$. Any solution (K_1,K_2) having $K_2>K_1$ must lie on the line $K_2=K_1+2\Phi^{-1}\left(\frac{Q+2}{2Q+2}\right)$. For simplicity in what follows x_0 will denote $2\Phi^{-1}\left(\frac{Q+2}{2Q+2}\right)$ and note that $x_0>0$.

Any intersection point (K_1,K_2) must simultaneously satisfy the equation $K_2=K_1+2\Phi^{-1}\left(\frac{Q+2}{2Q+2}\right)$ as well as the contour equation

$$1+\Phi(K_1-a)-\Phi(K_2-a)=\frac{Q}{1+Q}\ .$$

If we substitute the line into the contour equation we get $\Phi(K_1+x_0-a)-\Phi(K_1-a)=\frac{1}{1+Q}$. The unique[17] value of K_1 satisfying

[17] To prove uniqueness, note that $K_1=a-\frac{x_0}{2}$ is the point of maximum value of $\Phi(K_1+x_0-a)-\Phi(K_1-a)$.

this is $K_1 = a - \frac{x_0}{2}$. Since we must also satisfy $K_2 = K_1 + x_0$, we must have $K_2 = a + \frac{x_0}{2}$ and hence $K_1 + K_2 = 2a$. This means that provided $b > a$, $(\mathrm{K}_1, \mathrm{K}_2)$ is not in the region $2b \le K_1 + K_2$. Hence there is no intersection point in Region IV.

4.2 Characterising the Points $(\mathrm{K}_1, \mathrm{K}_2)$ Outside the Gorge

The points not lying on the gorge fall into two sets which are,

$$\left\{ (\mathrm{K}_1, \mathrm{K}_2) : \alpha^*(\mathrm{K}_1, \mathrm{K}_2) \; > \; \mathrm{Q}\beta^*(\mathrm{K}_1, \mathrm{K}_2) \right\}$$

and

$$\left\{ (\mathrm{K}_1, \mathrm{K}_2) : \alpha^*(\mathrm{K}_1, \mathrm{K}_2) \; < \; \mathrm{Q}\beta^*(\mathrm{K}_1, \mathrm{K}_2) \right\} .$$

In this section we will show that these two sets are just the connected regions **on either side of the gorge**. Before doing this we need to introduce some notation that will be used for that purpose.

Defining the Notation $K_1^G(K_1', K_2', \mathrm{Q})$

We consider an arbitrary point $(\mathrm{K}_1', \mathrm{K}_2')$ lying Region III or in Region IV. Consider a line through this point parallel to $K_1 + K_2 = a + b$. Such a line will be of the form $K_1 + K_2 = K_1' + K_2'$.

Now consider the gorge $\alpha^*(\mathrm{K}_1, \mathrm{K}_2) = \mathrm{Q}\beta^*(\mathrm{K}_1, \mathrm{K}_2)$. The point where $K_1 + K_2 = K_1' + K_2'$ meets the gorge will be denoted $K_1^G(\mathrm{K}_1', \mathrm{K}_2', \mathrm{Q})$ when it is useful to emphasise its dependence on the point $(\mathrm{K}_1', \mathrm{K}_2')$ and on the gorge parameter Q. More briefly it will be referred to simply as K_1^G.

The value K_1^G is the solution of the equation obtained from substituting $K_2 = K_1' + K_2' - K_1$ into the gorge equation.

Since the gorge equation is different in Region III and Region IV these are treated separately.

When (K_1', K_2') **lies in Region III i.e.** $a+b \le K_1' + K_2' \le 2b$;

The value K_1^G is the solution of

$$1+\Phi(K_1-a)-\Phi(K_1'+K_2'-K_1-a) = Q\,\Phi(K_1'+K_2'-K_1-b)-Q\Phi(K_1-b)$$

When (K_1', K_2') **lies in Region IV i.e.** $2b \le K_1' + K_2'$;

The value K_1^G is the solution of

$$1+\Phi(K_1-a)-\Phi(K_1'+K_2'-K_1-a) = -Q+2Q\Phi\left(\frac{K_1'+K_2'}{2}-K_1\right)$$

Notice also that because $K_1 = K_1^G$ and $K_2 = K_1' + K_2' - K_1^G$ are by definition the coordinates of a point on the gorge, we can also write $\hat{K}_2(K_1^G) = K_1' + K_2' - K_1^G$.

We will also introduce a notation $G(K_1)$, $H(K_1)$, $J(K_1)$ for the three functions of K_1 appearing in the above expressions. These are defined as

$$G(K_1) = 1+\Phi(K_1-a)-\Phi(K_1'+K_2'-K_1-a)$$

$$H(K_1) = Q\,\Phi\left(K_1' + K_2' - K_1 - b\right) - Q\Phi\left(K_1 - b\right)$$

$$\mathrm{J}(K_1) = -Q + 2Q\,\Phi\left(\frac{K_1' + K_2'}{2} - K_1\right)$$

Notice that:

$G(K_1)$ is monotone increasing in K_1

$H(K_1)$ is monotone decreasing in K_1

$\mathrm{J}(K_1)$ is monotone decreasing in K_1 .

and that $G(K_1') = \alpha^*(\mathrm{K}_1', \mathrm{K}_2')$;

$H(K_1') = Q\beta^*(\mathrm{K}_1', \mathrm{K}_2')$ in Region III

and $J(K_1') = Q\beta^*(\mathrm{K}_1', \mathrm{K}_2')$ in Region IV.

Note also that it is straightforward to show that for $(\mathrm{K}_1', \mathrm{K}_2')$ in Region III that $G(K_1^G) = \mathrm{H}(K_1^G)$, because $(K_1^G, \mathrm{K}_2' + \mathrm{K}_1' - K_1^G)$ is a point on the gorge and that for $(\mathrm{K}_1', \mathrm{K}_2')$ in Region IV that $G(K_1^G) = \mathrm{J}(K_1^G)$, again because $(K_1^G, \mathrm{K}_2' + \mathrm{K}_1' - K_1^G)$ is a point on the gorge.

4.2.1 Two Lemmae: We now prove for any point $(\mathrm{K}_1', \mathrm{K}_2')$ lying Region III or in Region IV that

Lemma 1 $(\mathrm{K}_1', \mathrm{K}_2') \in \left\{(\mathrm{K}_1, \mathrm{K}_2) : \alpha^*(\mathrm{K}_1, \mathrm{K}_2) > Q\beta^*(\mathrm{K}_1, \mathrm{K}_2)\right\}$

if and only if $\mathrm{K}_1' > K_1^G$

and

Lemma 2 $(K'_1, K'_2) \in \left\{ (K_1, K_2) \colon \alpha^*(K_1, K_2) < Q\beta^*(K_1, K_2) \right\}$
if and only if $K'_1 < K_1^G$

Proof of Lemma 1:
Proof when (K'_1, K'_2) **is in Region III:** Here we use the fact that $G(K_1)$ is monotone increasing, $H(K_1)$ monotone decreasing and $G(K_1^G) = H(K_1^G)$.

Now suppose $K'_1 > K_1^G$. Because of the monotone properties of $G(K_1)$ and $H(K_1)$, this gives $G(K'_1) > G(K_1^G) = H(K_1^G) > H(K'_1)$.
In Region III, $G(K'_1) > H(K'_1)$ is just $\alpha^*(K'_1, K'_2) > Q\beta^*(K'_1, K'_2)$.
Hence $(K'_1, K'_2) \in \left\{ (K_1, K_2) \colon \alpha^*(K_1, K_2) > Q\beta^*(K_1, K_2) \right\}$.

Conversely suppose that $G(K'_1) > H(K'_1)$ but assume that $K'_1 < K_1^G$.
Because of the monotone properties of $G(K_1)$ and $H(K_1)$, $K'_1 < K_1^G$ implies $G(K'_1) < G(K_1^G) = H(K_1^G) < H(K'_1)$ which is a contradiction.
Similarly if we suppose that $G(K'_1) > H(K'_1)$ but assume that $K'_1 = K_1^G$ we immediately have a contradiction with $G(K_1^G) = H(K_1^G)$.

Hence $G(K'_1) > H(K'_1)$ i.e. $\alpha^*(K'_1, K'_2) > Q\beta^*(K'_1, K'_2)$ implies $K'_1 > K_1^G$

Hence we have shown for (K'_1, K'_2) in Region III that $K'_1 > K_1^G$ if and only if

$$(K'_1, K'_2) \in \left\{ (K_1, K_2) \colon \alpha^*(K_1, K_2) > Q\beta^*(K_1, K_2) \right\}.$$

Proof when $\left(K'_1, K'_2\right)$ is in Region IV: Here we use the fact that $G\left(K_1\right)$ is monotone increasing, $J\left(K_1\right)$ is monotone decreasing and $G\left(K_1^G\right) = J\left(K_1^G\right)$.

The proof in Region IV is exactly analogous to the proof in Region III with the modification that in Region IV $J\left(K_1\right)$ plays the role played by $H\left(K_1\right)$ in Region III. With this substitution all the same steps of proof apply.

Hence we can show in analogous manner for $\left(K'_1, K'_2\right)$ in Region IV that $K'_1 > K_1^G$ if and only if

$$(K'_1, K'_2) \in \left\{\left(K_1, K_2\right): \alpha^*\left(K_1, K_2\right) > Q\beta^*\left(K_1, K_2\right)\right\}.$$

Proof of Lemma 2: In Lemma 2 we seek to show that

$$(K'_1, K'_2) \in \left\{\left(K_1, K_2\right): \alpha^*\left(K_1, K_2\right) < Q\beta^*\left(K_1, K_2\right)\right\}$$

if and only if $K'_1 < K_1^G$

The proof of Lemma 2 is exactly analogous to the proof of Lemma 1 and hence will not be given in detail here. The only difference is that we apply the condition $K'_1 < K_1^G$. Again we have to consider separately Region III (using $G\left(K_1\right)$ and $H\left(K_1\right)$) and Region IV (using $G\left(K_1\right)$ and $J\left(K_1\right)$). The monotone properties of $G\left(K_1\right)$, $H\left(K_1\right)$ and $J\left(K_1\right)$ plus $G\left(K_1^G\right) = H\left(K_1^G\right)$ in Region III and $G\left(K_1^G\right) = J\left(K_1^G\right)$ in Region IV, immediately provide the proof as in the case of Lemma 1.

Conclusion: The pictorial meaning of Lemma 1 is that $\left\{\left(K_1, K_2\right): \alpha^*\left(K_1, K_2\right) > Q\beta^*\left(K_1, K_2\right)\right\}$ is the region on the right hand side of the gorge and the pictorial meaning of Lemma 2 is that

$\left\{(K_1,K_2): \alpha^*(K_1,K_2) < Q\beta^*(K_1,K_2)\right\}$ is the region on the left hand side of the gorge as shown in Figure 4 below.

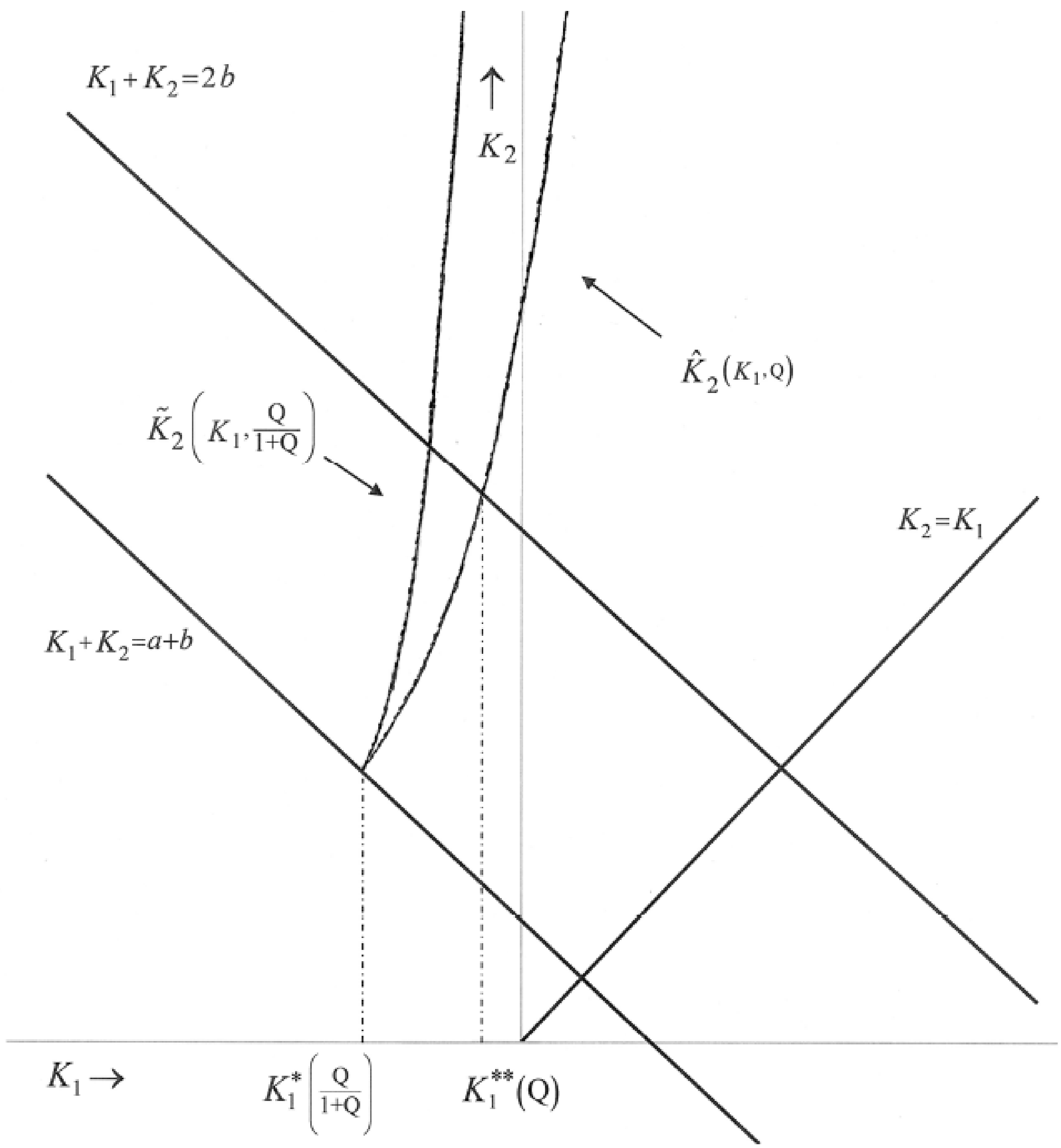


Figure 4: Graph of the "gorge" $\hat{K}_2(K_1,Q)$ for $K_1+K_2 \geq a+b$

4.2.2 A point not lying on the gorge cannot be a minimum (nor an inf)

The points not lying on the gorge fall into two sets which are,

$$\left\{ (K_1, K_2): \alpha^*(K_1, K_2) > Q\beta^*(K_1, K_2) \right\}$$

and

$$\left\{ (K_1, K_2): \alpha^*(K_1, K_2) < Q\beta^*(K_1, K_2) \right\}.$$

We have shown earlier that these two sets are just the regions **on either side of the gorge** which is an unbounded curve. We will now study the values of the function $\sup\left\{ \alpha^*(K_1, K_2), Q\beta^*(K_1, K_2) \right\}$ on either side of the gorge. On one side the value of this **sup** will be given by $\alpha^*(K_1, K_2)$ and on the other side the value is given by $Q\beta^*(K_1, K_2)$.

The important result to be shown below is that **for any point not lying on the gorge, there exists a point lying on the gorge and having a smaller value of** $\sup\left\{ \alpha^*(K_1, K_2), Q\beta^*(K_1, K_2) \right\}$.

Points in the Region $\left\{ (K_1, K_2): \alpha^*(K_1, K_2) > Q\beta^*(K_1, K_2) \right\}$

The value of the sup in the region is $\alpha^*(K_1, K_2)$. Let (K'_1, K'_2) be a point having $\alpha^*(K'_1, K'_2) > Q\beta^*(K'_1, K'_2)$. $K_1 + K_2 = K'_1 + K'_2$ is the line passing through (K'_1, K'_2) and parallel to $K_1 + K_2 = a+b$. As shown earlier in Lemma 1 of 4.2.1, moving from (K'_1, K'_2) along this line towards the gorge is a path along which K_1 is diminishing. The value $\alpha^*(K_1, K_2)$ along this line is given by substituting $K_2 = K'_1 + K'_2 - K_1$ into the equation

$$\alpha^*(K_1, K_2) = 1 + \Phi(K_1 - a) - \Phi(K_2 - a).$$

This gives

$$\lambda\left(K_1\right) = \alpha^*\left(K_1, K'_1 + K'_2 - K_1\right)$$
$$= 1 + \Phi\left(K_1 - a\right) - \Phi\left(K'_1 + K'_2 - K_1 - a\right)$$

This shows that $\lambda\left(K_1\right)$ is a monotone increasing function and hence that $\lambda\left(K_1\right)$ diminishes as K_1 moves along the line to the point of intersection with the gorge. This point of intersection was earlier defined as $K_1^G = K_1^G(K'_1, K'_2, Q)$ and hence we have, $\lambda\left(K'_1\right) > \lambda\left(K_1^G\right)$. It is straightforward to show that $\lambda\left(K'_1\right) = \alpha^*\left(K'_1, K'_2\right)$ and using the fact that $\left(K_1^G, K'_2 + K'_1 - K_1^G\right)$ is a point on the gorge it is straightforward to show that $\lambda\left(K_1^G\right) = \alpha^*\left(K_1^G, \hat{K}_2\left(K_1^G\right)\right)$. This shows that K_1^G is a point on the gorge having a value of $\alpha^*(K_1, K_2)$ strictly lower than the value $\alpha^*(K'_1, K'_2)$. QED.

Comment on Proof: Note that we saw earlier that the equation to find the numerical value of $K_1^G(K'_1, K'_2, Q)$ for a particular choice of K'_1, K'_2 and Q depended on whether $\left(K'_1, K'_2\right)$ is contained in Region III or in Region IV. But this fact is not overtly used in the above proof. This is because $\alpha^*(K_1, K_2)$ is given by the same expression in Regions III and IV and the result of Lemma 1 is derived using the fact that K_1^G depends on different equations in the two regions.

Points in the Region $\left\{\left(K_1, K_2\right): \alpha^*\left(K_1, K_2\right) < Q\beta^*\left(K_1, K_2\right)\right\}$

The value of the sup in this region is $Q\beta^*(K'_1, K'_2)$. Let $\left(K'_1, K'_2\right)$ be a point having $\alpha^*(K'_1, K'_2) < Q\beta^*(K'_1, K'_2)$. $K_1 + K_2 = K'_1 + K'_2$ is the line passing through $\left(K'_1, K'_2\right)$ and parallel to $K_1 + K_2 = a + b$. We have shown earlier that moving from $\left(K'_1, K'_2\right)$ along this line towards the gorge is a path along which K_1 is increasing. The value of $Q\beta^*(K_1,K_2)$ along this line is given by substituting $K_2 = K'_1 + K'_2 - K_1$ into the equations defining $\beta^*(K_1,K_2)$. This value will be denoted $\rho(K_1) = Q\beta^*(K_1, K'_1 + K'_2 - K_1)$ and will be evaluated separately for Regions III and IV. We will see in each case that the value of $\beta^*(K_1,K_2)$ along the line diminishes as the point moves towards the gorge.

In Region III: $a+b \le K_1 + K_2 \le 2b$, we have

$$\beta^*(K_1,K_2) = \Phi(K_2-b) - \Phi(K_1-b)$$

This gives

$$\begin{aligned}\rho(K_1) &= Q\beta^*(K_1, K'_1 + K'_2 - K_1) \\ &= Q\Phi(K'_1 + K'_2 - K_1 - b) - Q\Phi(K_1 - b)\end{aligned}$$

This defines $\rho(K_1)$ when (K'_1, K'_2) is in Region III.

In Region IV: $2b \le K_1 + K_2$, we have

$$\beta^*(K_1,K_2) = -1 + 2\Phi\left(\frac{K_2 - K_1}{2}\right) \quad .$$

This gives

$$\begin{aligned}\rho(K_1) &= Q\beta^*(K_1, K'_1 + K'_2 - K_1) \\ &= -Q + 2Q\Phi\left(\frac{K'_2 + K'_1}{2} - K_1\right)\end{aligned}$$

This defines $\rho(K_1)$ when (K'_1, K'_2) is in Region IV.

These equations for $\rho(K_1)$ show that in both cases, $\rho(K_1)$ is a monotone decreasing function and hence that $\rho(K_1)$ diminishes as K_1 moves along the line (K_1 increasing) to the point of intersection with the gorge. This point of intersection was earlier defined as $K_1^G = K_1^G(\mathrm{K}'_1, \mathrm{K}'_2, \mathrm{Q})$ and hence we have, $\rho\left(\mathrm{K}'_1\right) > \rho\left(K_1^G\right)$. It is straightforward to show that in both Region III and Region IV we have $\rho\left(\mathrm{K}'_1\right) = \mathrm{Q}\beta^*\left(\mathrm{K}'_1, K'_2\right)$ and since $\left(K_1^G, \mathrm{K}'_2 + \mathrm{K}'_1 - K_1^G\right)$ is a point on the gorge, it is straightforward to show in both regions that $\rho\left(K_1^G\right) = \mathrm{Q}\beta^*\left(K_1^G, \hat{K}_2\left(K_1^G\right)\right)$. This shows that K_1^G is a point on the gorge having value of $\mathrm{Q}\beta^*(\mathrm{K}_1, \mathrm{K}_2)$ strictly lower than the value $\mathrm{Q}\beta^*(\mathrm{K}'_1, \mathrm{K}'_2)$. QED.

Comment on Proof: Here, the fact that the equation for $\beta^*(\mathrm{K}_1, \mathrm{K}_2)$ depends on whether $(\mathrm{K}_1, \mathrm{K}_2)$ is in Region III or in Region IV, is used in the above proof to give the relevant expressions for $\beta^*\left(K_1, K'_1 + K'_2 - K_1\right)$.

Summary Conclusion

We have shown the important result that **for any point not lying on the gorge, there exists a point lying on the gorge which has a smaller value of** $\sup\left\{\alpha^*(\mathrm{K}_1, \mathrm{K}_2), \mathrm{Q}\beta^*(\mathrm{K}_1, \mathrm{K}_2)\right\}$.

Hence a point $(\mathrm{K}_1, \mathrm{K}_2)$ which is not on the gorge cannot be a minimum point of the function $\sup\left\{\alpha^*(\mathrm{K}_1, \mathrm{K}_2), \mathrm{Q}\beta^*(\mathrm{K}_1, \mathrm{K}_2)\right\}$.

This means that in looking for a point giving a minimum of

$$\sup\left\{\alpha^*(\mathrm{K}_1, \mathrm{K}_2), \mathrm{Q}\beta^*(\mathrm{K}_1, \mathrm{K}_2)\right\} ,$$

only the points on the gorge need to be considered.

4.3 The Gradient of the Implicit Function defining the Gorge

From the implicit function theorem, $\alpha^*(\mathrm{K}_1,\mathrm{K}_2) = \mathrm{Q}\beta^*(\mathrm{K}_1,\mathrm{K}_2)$ defines K_2 as an implicit function of K_1 for the values K_1 that are possible with the specific value of Q. The function values also depend on the value of Q. This function will be denoted $\hat{K}_2(K_1)$ for a set of K_1 values that has still to be identified.

In Region III: i.e. $a+b \le K_1 + K_2 \le 2b$, the gorge equation is

$$1+\Phi(K_1-a)-\Phi(K_2-a)-\mathrm{Q}\Phi(K_2-b)+\mathrm{Q}\Phi(K_1-b)=0$$

Applying the implicit function theorem to this gives,

$$\frac{d\hat{K}_2}{dK_1} = \frac{\Phi'(\mathrm{K}_1-a)+\mathrm{Q}\Phi'(\mathrm{K}_1-b)}{\Phi'(\hat{\mathrm{K}}_2-a)+\mathrm{Q}\Phi'(\hat{\mathrm{K}}_2-b)} > 0 \ ;\forall\, \mathrm{K}_1, \hat{\mathrm{K}}_2$$

where $\Phi'(u)$ denotes $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$.

$\frac{d\hat{K}_2}{dK_1}$ **On the Boundary** $K_1 + K_2 = a+b$

The Region III expression for $\frac{d\hat{K}_2}{dK_1}$ can be used to show that when $\left(K_1,\hat{K}_2(K_1)\right)$ lies on the line $K_1 + K_2 = a+b$, i.e. for points where the Region III gorge meets the lower boundary, we have

$$\frac{d\hat{K}_2}{dK_1} < 1 \quad ;\forall\, \mathrm{Q}>1$$

$$\frac{d\hat{K}_2}{dK_1} \;=\; 1 \qquad \text{when Q=0}$$

$$\frac{d\hat{K}_2}{dK_1} \;>\; 1 \qquad ;\forall\, \text{Q<1}$$

The proof uses the fact (proven in Section 4.1.1) that the intersection point of the gorge with the boundary $K_1 + K_2 = a+b$, is given by $K_1 = K_1^*\left(\frac{Q}{1+Q}\right) < \frac{a+b}{2}$ and uses the fact that $b > a$.

Comparing $\frac{d\hat{K}_2}{dK_1}$ **with** $\frac{d\tilde{K}_2}{dK_1}$ **:** We have seen in Sections 4.1.2 and 4.1.5 that the curves $\hat{K}_2(K_1)$ and $\tilde{K}_2(K_1)$ meet only at the point given by

$$K_1 = K_1^*\left(\frac{Q}{1+Q}\right) \text{ and } K_2 = a + b - K_1^*\left(\frac{Q}{1+Q}\right).$$

The values of the two gradients at this point can be compared by substituting $K_2 = a + b - K_1^*$ into the implicit function expressions for the two derivatives. It is straightforward to establish that $\frac{d\hat{K}_2}{dK_1} < \frac{d\tilde{K}_2}{dK_1}$ at the point $K_1 = K_1^*\left(\frac{Q}{1+Q}\right)$ and $K_2 = a + b - K_1^*\left(\frac{Q}{1+Q}\right)$.

The proof uses the fact that $K_1^*\left(\frac{Q}{1+Q}\right) < \frac{a+b}{2}$ and that $b > a$.

On the Boundary $K_1 + K_2 = 2b$

The Region III expression can be used to show that when $\left(K_1, \hat{K}_2(K_1)\right)$ lies on the line $K_1 + K_2 = 2b$, i.e. for points where the gorge meets the upper boundary of region III, we have for the Region III gradient

$$\frac{d\hat{K}_2}{dK_1} > 1$$

The proof uses the fact (see 4.1.3) that when $K_1^{**}(\mathrm{Q})$ is the K_1 value at which the gorge $\hat{K}_2(K_1)$ meets $K_1 + K_2 = 2b$, we have $K_1^{**}(\mathrm{Q}) < b$. We proved earlier (4.1.1) that the gorge has no intersection with $K_2 = K_1$. Hence the point having $K_1 = b$ and $K_2 = b$ cannot be a point on the gorge.

In Region IV: i.e. $2b \le K_1 + K_2$, the gorge equation is,

$$1 + \Phi(K_1 - a) - \Phi(K_2 - a) + \mathrm{Q} - 2\mathrm{Q}\,\Phi\left(\frac{K_2 - K_1}{2}\right) = 0$$

Applying the implicit function theorem to this, gives for all points $\left(K_1, \hat{K}_2(K_1)\right)$ on the gorge (even when $K_1 + \hat{K}_2 = 2b$),

$$\frac{d\hat{K}_2}{dK_1} = \frac{\Phi'\left(\mathrm{K}_1 - a\right) + \mathrm{Q}\,\Phi'\left(\frac{\hat{K}_2 - K_1}{2}\right)}{\Phi'\left(\hat{\mathrm{K}}_2 - a\right) + \mathrm{Q}\,\Phi'\left(\frac{\hat{K}_2 - K_1}{2}\right)} > 1\,;$$

It is straightforward to show this when $b > a$ using the fact that $K_1 + K_2 \ge 2b$ implies $K_1 + K_2 > 2a$ and using the fact that $\hat{K}_2(K_1) > K_1$ (Section 4.1.1).

The Gorge Gradient is continuous on the Boundary $K_1 + K_2 = 2b$

Substituting $\hat{K}_2 = 2b - K_1$ into

$$\frac{d\hat{K}_2}{dK_1} = \frac{\Phi'\left(\mathrm{K}_1 - a\right) + \mathrm{Q}\,\Phi'\left(\mathrm{K}_1 - b\right)}{\Phi'\left(\hat{\mathrm{K}}_2 - a\right) + \mathrm{Q}\,\Phi'\left(\hat{\mathrm{K}}_2 - b\right)} \quad \text{(gorge in Region III)}$$

and into

$$\frac{d\hat{K}_2}{dK_1} = \frac{\Phi'\left(\mathrm{K}_1 - a\right) + \mathrm{Q}\,\Phi'\left(\frac{\hat{K}_2 - K_1}{2}\right)}{\Phi'\left(\hat{\mathrm{K}}_2 - a\right) + \mathrm{Q}\,\Phi'\left(\frac{\hat{K}_2 - K_1}{2}\right)} \quad \text{(gorge in Region IV)}$$

the two resulting expressions are equal.

Both expressions reduce to $\dfrac{\Phi'\left(\mathrm{K}_1 - a\right) + \mathrm{Q}\,\Phi'\left(\mathrm{K}_1 - b\right)}{\Phi'\left(2b - K_1 - a\right) + \mathrm{Q}\,\Phi'\left(\mathrm{K}_1 - b\right)}$

4.4 The Value of $\sup\left\{ \alpha_d^* , \mathrm{Q}\,\beta_d^* \right\}$ along the gorge

By definition the gorge is $\left\{ \left(\mathrm{K}_1, \mathrm{K}_2\right) : \alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right) = \mathrm{Q}\,\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right) \right\}$.

For any point on the gorge, by definition the two elements $\alpha^*(\mathrm{K}_1, \mathrm{K}_2)$ and $\mathrm{Q}\,\beta^*(\mathrm{K}_1, \mathrm{K}_2)$ of the sup, are equal. Hence we have for any point $(\mathrm{K}_1, \mathrm{K}_2)$ on the gorge,

$$\sup\left\{ \alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right),\ \mathrm{Q}\,\beta^*\left(\mathrm{K}_1, \mathrm{K}_2\right) \right\} = \alpha^*(\mathrm{K}_1, \mathrm{K}_2)$$

We will now consider the value of $\alpha^*(\mathrm{K}_1,\mathrm{K}_2)$ as $(\mathrm{K}_1,\mathrm{K}_2)$ moves along the gorge starting from the boundary point $\left(K_1^*\left(\frac{\mathrm{Q}}{1+\mathrm{Q}}\right), a+b - K_1^*\left(\frac{\mathrm{Q}}{1+\mathrm{Q}}\right)\right)$.

The gorge is described by $K_1 \geq K_1^*\left(\frac{\mathrm{Q}}{1+\mathrm{Q}}\right)$ and $\hat{K}_2(K_1)$ which gives the K_2 coordinate of the gorge. The value of $\alpha^*(\mathrm{K}_1,\mathrm{K}_2)$ along the gorge is given by $H(K_1) = \alpha^*\left(\mathrm{K}_1,\hat{K}_2(K_1)\right)$.

The gradient of $H(K_1)$ is given by $\dfrac{\mathrm{dH}}{\mathrm{dK}_1} = \dfrac{\partial\alpha^*}{\partial \mathrm{K}_1} + \dfrac{\partial\alpha^*}{\partial \mathrm{K}_2}\times\left[\dfrac{d\hat{K}_2}{dK_1}\right]$

where $\dfrac{\partial\alpha^*}{\partial \mathrm{K}_1}$ and $\dfrac{\partial\alpha^*}{\partial \mathrm{K}_2}$ are to be evaluated at $\left(K_1, \hat{K}_2(K_1)\right)$. We desire to show that $\dfrac{\mathrm{dH}}{\mathrm{dK}_1} > 0$.

Note that whereas $\alpha^*(\mathrm{K}_1,\mathrm{K}_2)$ is given by the same expression in Regions III and IV, the gorge derivative $\dfrac{d\hat{K}_2}{dK_1}$ has different expressions in these two regions.

It was shown in Section 2.4 that throughout both Regions III and IV we have

$$\frac{\partial\alpha^*}{\partial \mathrm{K}_1} = \Phi'\left(\mathrm{K}_1 - a\right) \text{ and } \frac{\partial\alpha^*}{\partial \mathrm{K}_2} = -\Phi'\left(\mathrm{K}_2 - a\right).$$

where $\Phi'(u)$ denotes $\dfrac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}u^2}$.

In Region III we have,

$$\frac{d\hat{K}_2}{dK_1} = \frac{\Phi'\left(K_1-a\right)+Q\,\Phi'\left(K_1-b\right)}{\Phi'\left(\hat{K}_2-a\right)+Q\,\Phi'\left(\hat{K}_2-b\right)}$$

In Region IV we have,

$$\frac{d\hat{K}_2}{dK_1} = \frac{\Phi'\left(K_1-a\right)+Q\,\Phi'\left(\frac{\hat{K}_2-K_1}{2}\right)}{\Phi'\left(\hat{K}_2-a\right)+Q\,\Phi'\left(\frac{\hat{K}_2-K_1}{2}\right)}$$

Substituting these formulae into the definition of $\frac{dH}{dK_1}$ we get,

In Region III ,

$$\frac{dH}{dK_1} = \frac{Q\left[\Phi'\left(K_1-a\right)\Phi'\left(\hat{K}_2-b\right) - \Phi'\left(\hat{K}_2-a\right)\Phi'\left(K_1-b\right)\right]}{\Phi'\left(\hat{K}_2-a\right)+Q\,\Phi'\left(\hat{K}_2-b\right)}$$

Using the fact (Section 4.1.1) that for all points on the gorge we have $\hat{K}_2(K_1) > K_1$ and using $b > a$, it is straightforward to show that this expression satisfies,

$$\frac{dH}{dK_1} > 0 \ , \forall K_1, \hat{K}_2(K_1): \ K_1^{*b}\left(\frac{Q}{1+Q}\right) \le K_1 \le K_1^{**}(Q) \quad .$$

In Region IV,

$$\frac{\mathrm{dH}}{\mathrm{dK}_1} = \frac{\mathrm{Q}\Phi'\left(\frac{\hat{K}_2-K_1}{2}\right)\left[\Phi'\left(\mathrm{K}_1-a\right)-\Phi'\left(\hat{\mathrm{K}}_2-a\right)\right]}{\Phi'\left(\hat{\mathrm{K}}_2-a\right)+\mathrm{Q}\Phi'\left(\frac{\hat{K}_2-K_1}{2}\right)}$$

Using the fact that for all points on the gorge we have,

- $\hat{K}_2(K_1) > K_1$ (Section 4.1.1) and
- in Region IV, $K_1 + K_2 - 2a > 0$ because $K_1 + K_2 \geq 2b$ and $b > a$

, it is straightforward to show that $\Phi'\left(\mathrm{K}_1-a\right) > \Phi'\left(\hat{\mathrm{K}}_2-a\right)$ and hence that the Region IV expression also satisfies,

$$\frac{\mathrm{dH}}{\mathrm{dK}_1} > 0 \quad \forall K_1, \hat{K}_2(K_1): \quad K_1^{**}(\mathrm{Q}) \leq K_1 .$$

Note: By substituting $K_1 = K_1^{**}(\mathrm{Q})$ and $\hat{K}_2 = 2b - K_1^{**}(\mathrm{Q})$ into both expressions for $\frac{\mathrm{dH}}{\mathrm{dK}_1}$ it is straightforward to prove that both expressions give the same value i.e. $\frac{\mathrm{dH}}{\mathrm{dK}_1}$ is continuous at the boundary $K_1 + K_2 = 2b$.

Summary Conclusion

Hence we have shown that for all points $\left(K_1, \hat{K}_2(K_1)\right)$ on the gorge $\alpha^*(\mathrm{K}_1,\mathrm{K}_2) = \mathrm{Q}\beta^*(\mathrm{K}_1,\mathrm{K}_2)$, whether in Region III or in Region IV, we have $\frac{dH}{dK_1} > 0$. This means that the minimum value of $\alpha^*(K_1, K_2)$ on the gorge occurs at the boundary point $\left(K_1^*\left(\frac{Q}{1+Q}\right), a+b - K_1^*\left(\frac{Q}{1+Q}\right) \right)$ and this minimum value is given by $\alpha_0^{**}(Q) = \frac{Q}{1+Q}$. This is because the region of definition of $\hat{K}_2(K_1)$ is $K_1 \geq K_1^*\left(\frac{Q}{1+Q}\right)$.

In other words **if we confine ourselves to talking about gorge points**, the risk function $\sup\left\{\alpha^*(K_1, K_2),\, Q\beta^*(K_1, K_2)\right\}$ takes its minimum value on the gorge, at the boundary point $\left(K_1^*\left(\frac{Q}{1+Q}\right), a+b - K_1^*\left(\frac{Q}{1+Q}\right) \right)$ where $K_1^*\left(\frac{Q}{1+Q}\right)$ is the solution of $\Phi(K_1 - a) + \Phi(K_1 - b) = \frac{Q}{1+Q}$. At this point, the value of $\sup\left\{\alpha^*(K_1, K_2),\, Q\beta^*(K_1, K_2)\right\}$ is $\alpha_0^{**}(Q) = \frac{Q}{1+Q}$.

Major Conclusion

We have earlier shown that as far as looking for a minimum of $\sup\left\{\alpha^*(K_1, K_2),\, Q\beta^*(K_1, K_2)\right\}$ is concerned, only the points on the gorge need to be considered. A point which is not on the gorge cannot be a minimum of the function $\sup\left\{\alpha^*(K_1, K_2),\, Q\beta^*(K_1, K_2)\right\}$.

Now we have shown that **considering only gorge points**, the risk function $\sup\left\{\alpha^*(\mathrm{K}_1,\mathrm{K}_2)\,,\,\mathrm{Q}\,\beta^*(\mathrm{K}_1,\mathrm{K}_2)\right\}$ takes its minimum value on the gorge at the boundary point $\left(K_1^*\,,a+b-K_1^*\right)$ where $K_1^*\left(\frac{\mathrm{Q}}{1+\mathrm{Q}}\right)$ is the solution of $\Phi\left(K_1-a\right)+\Phi\left(K_1-b\right)=\frac{\mathrm{Q}}{1+\mathrm{Q}}$. At this point the value of $\sup\left\{\alpha^*(\mathrm{K}_1,\mathrm{K}_2)\,,\,\mathrm{Q}\,\beta^*(\mathrm{K}_1,\mathrm{K}_2)\right\}$ is $\alpha_0^{**}(\mathrm{Q})=\frac{\mathrm{Q}}{1+\mathrm{Q}}$.

The objective posed in Section 1 of this report was to find the decision rule **d**: IR → {**AH₀** , **RH₀**} to minimize $\sup\left\{\alpha^*(\mathrm{K}_1,\mathrm{K}_2)\,,\,\mathrm{Q}\,\beta^*(\mathrm{K}_1,\mathrm{K}_2)\right\}$ where $K_1 \le K_2$ and the interval $\left[K_1\sigma\,,K_2\sigma\right]$ is the acceptance region of **d**(x). We have now shown that the minimum value with respect to $\left(\mathrm{K}_1,\mathrm{K}_2\right)$ of the function $\sup\left\{\alpha^*(\mathrm{K}_1,\mathrm{K}_2)\,,\,\mathrm{Q}\,\beta^*(\mathrm{K}_1,\mathrm{K}_2)\right\}$ is $\frac{\mathrm{Q}}{1+\mathrm{Q}}$ and that $\left(K_1^*\left(\frac{\mathrm{Q}}{1+\mathrm{Q}}\right),a+b-K_1^*\left(\frac{\mathrm{Q}}{1+\mathrm{Q}}\right)\right)$ is the unique minimum point. This unique minimum point defines the minimax rule for the specific value of Q.

The results obtained also show of course that the optimal choice of K_1 and K_2 will satisfy $K_1+K_2=a+b$. In other words the interval $\left[K_1\sigma\,,K_2\sigma\right]$ of the minimax rule is always symmetrically placed relative to the interval $\left[a\sigma\,,b\sigma\right]$. This is not surprising given the intrinsic symmetry of the problem.

5. Use of Minimax Tests for Material Balance Assessment

In this section we look at the practical application of the minimax tests. These can be applied for assessing the acceptability or not of an incomplete nuclear material balance or of a material balance where some facility measurement methods may be creating small measurement biases. This assessment can be carried out by the facility management or by an external control authority. Each minimax test compares the balance value to acceptance limits $\left[K_1, K_2\right]$ that are determined from,

5) the standard deviation (σ_{MUF}) of the balance (complete or incomplete). This σ_{MUF} represents the cumulative contribution of those measurement error components that have probability distributions. Computing σ_{MUF} requires knowing the measurement history of the material that has been processed during the balance period and to which the accounts refer, as well as knowing the probability distributions of the relevant measurement errors. In this report we consider that σ_{MUF} is a known value. The methods used to compute σ_{MUF} for practical situations are described in the report[18].

6) the management or inspector's aversion to false alarms. An alarm is the trigger for an investigation to determine the cause of the alarm. Aversion is reflected in a desire to choose the acceptance limits so that false alarms have low probability. This aversion is represented as a requirement to choose the statistical test rule so that the maximum false alarm probability is equal to a desired target value. The acceptable maximum false alarm probability of the inspector has earlier been denoted α_0. Section 2 of this report has derived the expressions for the maximum false alarm probability of any inspector decision rule. These expressions were denoted $\alpha^*\left(\mathrm{K}_1, \mathrm{K}_2\right)$.

---

[18] Statistical Models for σ_{MUF} Computations (Material Balance Testing). M. T. Franklin, NUSEC, IPSC, DG JRC, September 2008.

7) Upper and lower tolerance limits for the true amounts of any hold-up that has not been included in the balance. Testing with a composite null hypothesis can be applied to the problem of assessing the acceptability of a material balance when some small amount of well identified material has not been measured. Unmeasured hold-up is not included in the balance computation and hence the balance is an incomplete balance. This incomplete balance is the sum of a mean value (determined by the hold-up that is not accounted for) plus an accumulation of measurement error incorporated in the accounting values of the material that has been included in the balance. The term tolerance limit implies that the facility management or the control authority consider that holdup amounts have not been included in the balance and that possible effects in this range must be allowed for in assessing the material balance.

8) Upper and lower tolerance limits for the values of any uncorrected biases in the facility measurement process. Some facility measurement methods may create small measurement biases even though the mass values have been generated by correctly applied procedures. The measurement specialists can be aware that such biases can exist and that their existence is difficult to establish. They may consider that the possibility of a small bias should be allowed for in assessing the material balance.

The factors (1) and (2) reflect respectively the intrinsic measurement uncertainty in the balance (σ_{MUF}) and the attitude to risk of false alarms (α_0). The factors (3) and (4) reflect the completeness or not of the accounts and the perceived need to allow for uncorrected bias in accounting measurements. The factors (3) and (4) are used to generate the parameters **a** and **b** of the composite null hypothesis $\left[a\,\sigma^*_{MUF}\,,\ b\,\sigma^*_{MUF}\right]$. Establishing these values for **a** and **b,** is the subject of sections 5.2 (incomplete balance) and 5.3 (possible measurement bias).

5.1 The Basic Model of a Nuclear Material Balance

In this section, we introduce a notation used to describe nuclear materials accountancy. This notation provides a general equation that is referred to as the material balance identity. This equation underlies any analysis of the acceptability of a set of accounts. The material balance is defined as,

Balance = Beginning Inventory + Receipts - Shipments - Ending Inventory

The balance is also referred too as "material unaccounted for" or "MUF" and is sometimes written,

$$\mathbf{MUF = B\,I + R - S - E\,I\ .}$$

By basic model is meant a formal description of a set of accounts consisting of item mass values that provide a material balance (MUF) and that are linked to a set of real existing items. We have the following definitions:

N_i is the number of items referred to in the accounts of the i^{th} MUF component (BI, R, S, EI; e.g. i=1 means BI, etc.).

M_{ik} represents the **true mass**[19] of the k^{th} item in the i^{th} MUF component $i=1,2,3,4$; $k=1,2,\ldots,N_i$;

Z_{ik} represents the accounting mass value for the k^{th} item in the i^{th} MUF component ;

$L_{ik} = Z_{ik}-M_{ik}$ is the **accounting discrepancy**[20] for the k^{th} item in the i^{th} MUF component ;

Causes of Discrepancies: Discrepancies will always be non-zero because even when good practice procedures are correctly applied, the accountancy

[19] Note that the **true** mass can never be known. It can only be measured.

[20] The term discrepancy is used for this concept because it is traditional practice to keep the term "difference" for the difference between accounting value **and inspector measured value**. When accountancy is carried out correctly, all discrepancies are legitimate measurement errors.

mass will incorporate intrinsic measurement errors. Intrinsic measurement errors are a type of discrepancy whose probability distributions are usually well known because the measurement methods have usually been carefully studied. There can however be other causes of accounting discrepancies. For example, discrepancies can be caused by **unintended human errors**. Discrepancies caused by unintended human errors could be considered as having a probability distribution, but it may be quite difficult to define how this distribution could be estimated. Legitimate measurement errors and unintended human errors are not the only causes that must be considered. For verification purposes, we may have to consider the hypothesis that discrepancies could be part of a **deliberate strategy to falsify** the accounts. In this situation, the idea of a probability distribution for the discrepancies may be quite inappropriate.

Because we shall be considering different kinds of discrepancy, we do not at this point make any assumptions characterising the discrepancies $L_{\mathbf{ik}}$. As mentioned discrepancies can be caused by,

- legitimate measurement errors. This may include uncorrected biases[21] as one of the forms of measurement error contribution.
- inadvertent human errors e.g. errors in data recording for operating records or errors in data processing of operating records; or errors in performance of measurement procedures.
- falsification of accounts or records (by state, by facility management, by criminal insiders).

We now go on to make some fundamental definitions in terms of discrepancies and the MUF equation. The facility MUF is defined by the accountancy material balance equation as:

$$MUF = \sum_{i=1}^{4} \operatorname{sgn}(i) \sum_{k=1}^{N_i} Z_{ik}$$

Using the definition of discrepancy we can write this as,

[21] A bias is a measurement error contribution which does not have a probability distribution.

$$MUF = \sum_{i=1}^{4} \operatorname{sgn}(i) \sum_{k=1}^{N_i} \left(M_{ik} + L_{ik} \right)$$

This can also be written as:

$$MUF = \sum_{i=1}^{4} \operatorname{sgn}(i) \sum_{k=1}^{N_i} M_{ik} + \sum_{i=1}^{4} \operatorname{sgn}(i) \sum_{k=1}^{N_i} L_{ik}$$

where the first term (which contains only the true mass values), is the true material balance and will be denoted $\mathrm{MUF}_{\mathbf{TRUE}}$,

$$MUF_{TRUE} = \sum_{i=1}^{4} \operatorname{sgn}(i) \sum_{k=1}^{N_i} M_{ik}$$

Note that we **do not** assume that $\mathrm{MUF}_{\mathbf{TRUE}}$ is zero. Of course, if all material in each balance component is accounted for, it will be zero.

We referred earlier to the assumption that the set of items referred to in each MUF component $\{Z_{\mathbf{ik}}$, i=1,2,3,4; k=1,2,.. .,$N_{\mathbf{i}}$ $\}$, **are a set of identifiable objects whose existence as objects can be verified by an inspector.** The declared values $\{Z_{ik}\}$ in the accounts determine MUF. Only when the set of declared items exist and hence the true amounts in items are fixed, are the individual discrepancies $\{L_{ik}\}$ determined. By the way in which discrepancy is defined, MUF, $\mathrm{MUF}_{\mathbf{TRUE}}$ and the $\{L_{ik}\}$ will always satisfy the balance identity.

Note however that it is not assumed a priori that these objects actually contain nuclear material – verifying this is part of the purpose of verification. The analysis of remeasurement results carried out for verification along with the accounting data, allows us to make statements about these sets of physical objects and the related accounts. These statements can be of two kinds. The first kind, are statements that the accounts are coherent in themselves as a set of documents (i.e the balance is acceptable). The second kind are statements that the accounts agree with the reality of the material really present in the physical set of objects available for verification (values resulting from independent re-measurement give acceptable differences from the accounting values). We can make statements of the first kind for any set of documents pretending to be a set of accounts. We cannot however make statements of the second kind if we do not have a designated physical set of objects that can be verified. Hence we see that sets of real objects must be presented as the objects referred to by the accounts. We say that the four sets of material form a **closed material balance** if and only if $MUF_{\mathbf{TRUE}}$ is zero.

Note also that **in the basic model,** there is no assumption that the true MUF is actually zero for the sets of material being referred to. We do not assume that these sets of material represent a true material balance equal to zero.

Note that one could have a set of declared values and a satisfactory MUF all of which are a pure fiction. It is the linking to the set of existing objects (which the accounts are declared to be describing), which allows us to define the $\{L_{ik}\}$ and allows us to detect missing material or unacceptable accounts.

The Total Discrepancy: The second term in the MUF equation represents the accumulated discrepancy in the accounts. This will be denoted L_{MUF} where

$$L_{MUF} = \sum_{i=1}^{4} \operatorname{sgn}(i) \sum_{k=1}^{N_i} L_{ik}$$

This notation can be used with the earlier definition of $MUF_{\mathbf{TRUE}}$ to write,

$$MUF = MUF_{TRUE} + L_{MUF}$$

This equation is of fundamental importance and is sometimes referred to as the **material balance identity**.

5.2 Tolerance Limits for the Balance Effect of Hold-up Amounts

Here we consider the problem of assessing the acceptability of a material balance when some small amount of well identified material has not been measured. In such cases we have a well defined location containing material at both BI and at EI and the material is difficult to measure.

Suppose we let $\mathrm{M}_{\text{BI-holdup}}$ and $\mathrm{M}_{\text{EI-holdup}}$ represent respectively the true masses of the material unaccounted for in BI and EI. Suppose that for evaluation purposes, we construct a balance ignoring the two hold-up amounts. We write the balance identity for the incomplete balance as:

$$MUF^{*} = \sum_{i=1}^{4} \operatorname{sgn}(i) \sum_{k=1}^{N_i} M_{ik} + \sum_{i=1}^{4} \operatorname{sgn}(i) \sum_{k=1}^{N_i} L_{ik}$$

Where the summation excludes the hold-up in BI and EI and * denotes the incomplete nature of the accounts. We also have the compact version,

$$MUF^{*} = MUF^{*}_{TRUE} + L^{*}_{MUF}$$

Both $\mathrm{M}_{\text{BI-holdup}}$ and $\mathrm{M}_{\text{EI-holdup}}$ are left out of MUF^{*}_{TRUE} . If they were included (completing the sets of material for both BI and EI) it would give a $\mathrm{MUF}_{\mathrm{TRUE}} = 0$. Hence we have,

$$MUF^{*}_{TRUE} + M_{BI,holdup} - M_{EI,holdup} = 0$$

We then have,

$$MUF^{*}_{TRUE} = M_{EI,holdup} - M_{BI,holdup}$$

In many cases it is possible to establish upper and lower tolerance limits for the true amount of material in hold-up. Such limits are not an estimate of what is contained in the hold-up instead they are a statement of what values are to be tolerated. These limits can be derived taking account of the processing history and technology generating the hold-up. Process knowledge and history may allow us to choose upper and lower bounds for each of $\mathrm{M}_{\text{BI-holdup}}$ and $\mathrm{M}_{\text{EI-holdup}}$.

If we have upper and lower bounds for what are acceptable values for each of $\mathrm{M}_{\text{BI-holdup}}$ and $\mathrm{M}_{\text{EI-holdup}}$, we can then define,

$$\text{Upper bound of } \mathrm{MUF}^{*}_{\mathrm{TRUE}} = \text{Upper bound of } M_{EI,holdup} - \text{Lower bound of } M_{BI,holdup}$$

and similarly,

$$\text{Lower bound of } \mathrm{MUF}^{*}_{\mathrm{TRUE}} = \text{Lower bound of } M_{EI,holdup} - \text{Upper bound of } M_{BI,holdup}$$

If we have upper and lower bounds for $\mathrm{MUF}^{*}_{\mathrm{TRUE}}$, the incomplete balance MUF^{*} is the sum of a mean value contained between the upper and lower bounds plus an accumulation of measurement error intrinsic to the accounting values of the material that has been measured. In sections 2-5, these upper and lower bounds for MUF^{*}_{TRUE} were denoted by the interval $\left[a\,\sigma^{*}_{MUF},\ b\,\sigma^{*}_{MUF}\right]$ where **a** and **b** may be positive or negative[22]. From the point of assessing the accounts, we will now consider that a situation having $MUF^{*}_{TRUE} \in \left[a\,\sigma^{*}_{MUF},\ b\,\sigma^{*}_{MUF}\right]$ is a situation to be tolerated.

As far as any statistical test of the accounts is concerned, the condition: $MUF^{*}_{TRUE} \in \left[a\,\sigma^{*}_{MUF},\ b\,\sigma^{*}_{MUF}\right]$ now becomes the **composite null hypothesis** for the test.

[22] The upper and lower bounds are in mass units and have no relationship to standard deviations. It is convenient however to express them in units of σ^{*}_{MUF} as this will make the statistical test formulae look simpler.

5.3 Tolerance Limits for the Balance Effect of Uncorrected Bias

Here we consider a set of accounts with biases in some measurements. We return again to the balance identity,

$$MUF = \sum_{i=1}^{4} \operatorname{sgn}(i) \sum_{k=1}^{N_i} M_{ik} + \sum_{i=1}^{4} \operatorname{sgn}(i) \sum_{k=1}^{N_i} L_{ik}$$

We can assume that all material has been accounted for and hence that MUF_{TRUE} is zero or we can assume that we are dealing with an incomplete balance MUF^*. Now however we make different assumptions about the measurement system. We assume that the mass values have been generated by correctly applied procedures but that some measurement methods create small measurement biases. In other words, the mean of some of the measurement errors is non-zero and some measured mass values incorporate such biases.

This means that some L_{ik} are made up of two components. One component will remain constant even if the facility measurement were repeated, this component does not have a probability distribution and is called a **measurement bias**. The second component is a measurement variation whose value would change if the measurement (or its calibration) were repeated. This component has a probability distribution which is taken account of in the computation of σ_{MUF} . The mean of this component is zero.[23]

To make a formal representation of measurement bias we introduce the notation

$$L_{ik} = B_{ik} + \eta_{ik}$$

where B_{ik} denotes the bias component of L_{ik} and η_{ik} denotes the

[23] This is implicit in the definition of bias. Bias is the mean of the measurement error.

component of L_{ik} having a probability distribution. The accumulated discrepancy is L_{MUF} where,

$$L_{MUF} = \sum_{i=1}^{4} \operatorname{sgn}(i) \sum_{k=1}^{N_i} L_{ik} = \sum_{i=1}^{4} \operatorname{sgn}(i) \sum_{k=1}^{N_i} \left(B_{ik} + \eta_{ik}\right)$$

And this can be written,

$$L_{MUF} = B_{MUF} + \eta_{MUF}$$

The material balance identity can now be written

$$MUF = MUF_{TRUE} + B_{MUF} + \eta_{MUF}$$

where B_{MUF} is the net effect of bias and η_{MUF} is the cumulative error component having a probability distribution with zero mean.

In this situation σ_{MUF} is the same thing as σ_{η} and the mean value of MUF in this case is B_{MUF} because we are considering that a complete balance of material is being discussed (all material is accounted for).

If bias exists and if MUF is tested with a null hypothesis $MUF_{TRUE} = 0$, the non-zero bias will give higher risk of alarm than the target false alarm probability chosen for the test. This may mean that the material balance will be suspected whereas the real explanation is that there is some measurement bias. Alternatively if bias is suspected but not integrated into the statistical test, test alarms may be ignored as being “probably due to bias” without analytical support for this conclusion. Avoiding these types of situation is the motivation for having a composite null hypothesis that makes allowance for some tolerable bias.

To allow for possible bias in the measurement process, it is necessary to provide lower and upper limits for tolerated values of B_{MUF}. Given such limits, the testing of the complete balance can be carried out using a composite null hypothesis. As an example, we can consider the case of a single potential source of bias. Suppose that the source of bias refers to material that is input to the process. Suppose that the accountancy value for the total amount of such material in BI and R is Z kgs. Suppose that the measurement specialist considers that the measurement method can have a bias per unit mass with a value between λ_1 Kgs and λ_2 Kgs per unit mass.

This can be interpreted as saying that the accountancy value Z may suffer from a bias contribution $\lambda \, M$ where $\lambda \in [\lambda_1, \lambda_2]$ and where M is the true value of the material referred to by Z. The bias contribution can be estimated therefore by the lower and upper limits $\frac{\lambda_1 Z}{1+\lambda_1}$ Kg and $\frac{\lambda_2 Z}{1+\lambda_2}$. In the simple case where there is only one source of bias being considered, these are then the limits for tolerated values of B_{MUF}. For application in the formula for minimax tests, these limits must be used to generate the values **a** and **b** for the composite null hypothesis in the notation $\left[a\sigma^*_{MUF} \, , \, b\,\sigma^*_{MUF} \right]$.

Note that this is a very simple example for a measurement system in which the bias contribution to Z is a linear function of the true value M i.e. of the form λM and hence the measured value Z can be used as above i.e. $\frac{\lambda \; Z}{1+\lambda}$ to estimate $\lambda \; M$.

Many situations will be more complex than this very simple case. Any accountancy mass value Z will be a product of bulk determinations (e.g. volume measurement), metal factors (U or Pu) and in the case of ^{235}U, an enrichment determination. Each of the measurement methods used for bulk, metal concentration and enrichment, will have its own potential sources of bias, of which only the specialists in that method can judge the potential magnitude. A potential bias in one of these methods will transmit into the accounts in function of which particular mass values in the accounts have been determined using that method. When a bias is judged of potential relevance, the accountant will have to propagate its effect in the accounts to determine the range of its potential contribution to B_{MUF}. In a complex application, there may be several relevant sources of potential bias. In that case, each source of bias will generate a range of possible contribution to B_{MUF}. Taking account of all of these, the global limits for B_{MUF} will be computed exactly as when there were several hold-up contributions. In other words the global lower limit will be computed from the combination of source values that gives the smallest lower limit. Similarly the global upper limit will be computed from the combination of source values that gives the largest lower limit.

The most General Situation: As we seen above, a composite null hypothesis might be used in two kinds of situation,

- when there is unmeasured hold-up which is ignored in computing the balance but we are able to establish a range of hold-up true values that could occur,

- when all material has a correctly measured value in the balance accounts but there may be uncorrected biases and where it is possible to establish a range of values that can be tolerated for the cumulative effect of bias in the material balance.

Sometimes a composite null hypothesis for decision making can be a combined effect of both types of contribution. In this case the upper and lower bounds for the test are the combination of the separately determined upper and lower bounds for B_{MUF} and $\mathrm{MUF}_{\mathrm{TRUE}}$.

If the composite null hypothesis were true, the balance (complete or incomplete) is the sum of a mean value contained between upper and lower bounds plus an accumulation of measurement error intrinsic to the accounting values of the material that has been measured. As before, if we know the probability distributions of the measurement error components which have probability distributions, the standard deviation of such measurement error in MUF can be computed and will be denoted σ_{MUF}. This value is determined from the measurement history of the material that has been processed during the balance period along with the probability distributions of the respective measurement errors. In the most general application, the upper and lower bounds that are the composite null hypothesis are derived from a combination of the separately determined upper and lower tolerance bounds for B_{MUF} and $\mathrm{MUF}_{\mathrm{TRUE}}$.

5.4 Carrying out the Minimax Test

Section 4 derived the minimax decision rule or test for specific values of p_1, p_2, a, b and $\sigma = \sigma_{MUF}$. The minimax test is described by an acceptance interval $\left[K_1\sigma, K_2\sigma\right]$ symmetric about the null hypothesis $\left[a\sigma, b\sigma\right]$ and characterised by the equation

$$\Phi\left(K_1 - a\right) + \Phi\left(K_1 - b\right) = \alpha_0$$

$$\text{where } \alpha_0 = \frac{p_2}{p_1 + p_2}.$$

This equation[24] is solved for $K_1^* = K_1^*\left(\alpha_0\right)$ and the acceptance interval for the material balance is given by $\left[K_1^*\sigma, \left(a+b-K_1^*\right)\sigma\right]$. The next section provides tables of numerical values of $K_1^*\left(\alpha_0\right)$ computed using the above formulae.

Section 4 also showed that $\alpha_0 = \dfrac{p_2}{p_1 + p_2}$ was the maximum false alarm probability $\alpha^*\left(K_1^*, a+b-K_1^*\right)$ for that minimax rule.

[24] Where Φ is the standardised Gaussian distribution function

$$\Phi\left(K\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{K} e^{-\frac{1}{2}u^2} du$$

In what follows we will assume that the inspector has a probability value α_0 which is the **largest** false alarm probability that he wishes to tolerate. He wishes therefore to choose a minimax test, whose largest false alarm probability over the range of null hypothesis values $\left[a\,\sigma^*_{MUF} ,\ b\,\sigma^*_{MUF} \right]$, will be α_0 . He will solve the above equation for $K_1^*(\alpha_0)$ and will then use a decision rule of the form,

"Accept the balance only if
$$K_1^*\sigma^*_{MUF} \le MUF^* \le \left(a+b-K_1^*\right)\sigma^*_{MUF} \quad \text{"}$$

Note that $K_1^*(\alpha_0)$ can be positive or negative.

Different notations describing the same decision rule. Different notations are possible for describing this minimax test. An alternative notation is to write the test in the form,

"Accept the balance only if $\dfrac{a+b}{2} - K \le \dfrac{MUF^*}{\sigma^*_{MUF}} \le \dfrac{a+b}{2} + K$ "

In this notation, K has to be chosen (K > 0) by solving the equation,

$$\alpha_0 = 2 - \Phi\left(K + \frac{b-a}{2}\right) - \Phi\left(K - \frac{b-a}{2}\right)$$

A **third possible notation** is to write the minimax test in the form

"Accept the balance only if $a - K \le \dfrac{MUF^*}{\sigma^*_{MUF}} \le b + K$ "

In this notation the equation for computing K becomes

$$\alpha_0 = 2 - \Phi\left(K + b - a\right) - \Phi\left(K\right)$$

From the form of the equation, K can be positive or negative though usually it will be positive because α_0 is small.

5.5 Numerical Solutions and their Properties

The traditional approach to testing a material balance is to test against the simple null hypothesis that the mean of the Gaussian distribution is zero. This is equivalent to saying that the true balance is zero and all measurement methods are without bias. This takes no account of the possible existence of hold-up or uncorrected bias. For this reason, the traditional test can be seen as failing to address the real problem facing facility operators. This can produce situations where the traditional test gives a significant value but this result is assumed to be a spurious alarm due to bias or hold-up or whatever. The composite hypothesis approach provides a practical way of integrating quantitative hypotheses about hold-up and bias into the statistical procedure. Applied correctly, it leaves little leeway for delegitimating the statistical evaluation when the balance test gives an alarm.

In this section we look at examples of the acceptance regions generated by the minimax tests when applied to composite hypotheses. In doing this, comparison with the simple null hypothesis test is instructive. In the notation we have used for composite hypotheses, the traditional simple hypothesis $MUF_{TRUE} = 0$, is represented by having $a = b = 0$ in the specification of the composite hypothesis $\left[a\sigma, b\sigma\right]$. When $a = b = 0$, the minimax method of solving $\alpha_0 = \Phi\left(K^* - b\right) + \Phi\left(K^* - a\right)$ for $K_1^* < \frac{a+b}{2}$ and putting $K_2 = a + b - K_1^*$, is equivalent to solving $\alpha_0 = 2\left[1 - \Phi\left(K\right)\right]$ for $K > 0$ and using the traditional rule,

“Accept the balance only if $-K \le \dfrac{MUF}{\sigma_{MUF}} \le +K$ ”.

Note that $-K$ which is the analogue of K_1^* will satisfy $K = -\Phi^{-1}\left(\frac{\alpha_0}{2}\right)$.

Values of K_1^* and $K_2^*\left(K_1^*\right) = a + b - K_1^*$ for different values of b and α_0^* but all computed with $a = 0$.

We now look at balance acceptance rules for composite null hypotheses for different values of a, b and α_0. For simplicity we set $a = 0$ and then $b > 0$ also represents the range $b - a$ of the null hypothesis. The values of K_1^* and $K_2^*\left(K_1^*\right) = a + b - K_1^*$ in these tables are computed by solving $\alpha_0^* = \Phi\left(K - b\right) + \Phi\left(K\right)$ for K_1^* and then putting $K_2 = a + b - K_1^*$.

Table 1.1 $\alpha_0^* = 0.1$						
Value of 'b' →	0,0000	0,2500	0,5000	1,0000	2,0000	3,0000
$K_1^* \rightarrow$	-1,6449	-1,5327	-1,4456	-1,3388	-1,2845	-1,2816
$K_2^* = a + b - K_1^*$	1,6449	1,7827	1,9456	2,3388	3,2845	4,2816

Table 1.2 $\alpha_0^* = 0.05$						
Value of 'b' →	0,0000	0,2500	0,5000	1,0000	2,0000	3,0000
$K_1^* \rightarrow$	-1,9600	-1,8502	-1,7697	-1,6815	-1,6461	-1,6449
$K_2^* = a + b - K_1^*$	1,9600	2,1002	2,2697	2,6815	3,6461	4,6449

Table 1.3 $\alpha_0^* = 0.02$						
Value of 'b' →	0,0000	0,2500	0,5000	1,0000	2,0000	3,0000
$K_1^* \rightarrow$	-2,3263	-2,2193	-2,1462	-2,0759	-2,0543	-2,0538
$K_2^* = a + b - K_1^*$	2,3263	2,4693	2,6462	3,0759	4,0543	5,0538

Table 1.4 $\alpha_0^* = 0.01$						
Value of 'b' →	0,0000	0,2500	0,5000	1,0000	2,0000	3,0000
$K_1^* \rightarrow$	-2,5758	-2,4707	-2,4022	-2,3422	-2,3266	-2,3263
$K_2^* = a + b - K_1^*$	2,5758	2,7207	2,9022	3,3422	4,3266	5,3263

Table 1.5 $\alpha_0^* = 0.005$						
Value of 'b' →	0,0000	0,2500	0,5000	1,0000	2,0000	3,0000
$K_1^* \rightarrow$	-2,8070	-2,7036	-2,6393	-2,5876	-2,5760	-2,5758
$K_2^* = a + b - K_1^*$	2,8070	2,9536	3,1393	3,5876	4,5760	5,5758

Looking at Tables 1.1 – 1.5, notice that when $b = 0.0$ (i.e. the simple null hypothesis), the values of K_1^* and $K_2^* = a + b - K_1^*$ are those given by $K_1^*(\alpha_0) = \Phi^{-1}\left(\frac{\alpha_0}{2}\right)$. As mentioned earlier, they correspond to the traditional two sided test values from the standardised Gaussian tables.

In these tables we also see the values of K_1^* illustrating the properties described in Section 2.5.1 for $\alpha_0 < \frac{1}{2}$. In any of the tables, we have $K_1^* < a = 0$ and as b gets larger, the absolute value of K_1^* is decreasing. We proved in Section 2.5.1, that the larger b becomes, the more $\left|K_1^*(\alpha_0) - a\right|$ is contracting. Since $a = 0$ the value K_1^* is tending to $\Phi^{-1}(\alpha_0)$.
In 2.5.1, we showed that $\lim_{b \to \infty} K_1^*(\alpha_0, a, b) = a + \Phi^{-1}(\alpha_0)$.

Detection Power of the Minimax Test

In Section 2 we saw that for any test given by $\left(\mathrm{K}_1, \mathrm{K}_2\right)$,

$$\alpha\left(\theta, \mathrm{K}_1, \mathrm{K}_2\right) = 2 - \Phi\left(-K_1 + \frac{\theta}{\sigma}\right) - \Phi\left(K_2 - \frac{\theta}{\sigma}\right)$$

is the probability of rejecting H_0 given θ.

For a symmetric test of size α_0, we have $K_1 = K_1^*\left(\alpha_0\right)$ and $K_2 = a + b - K_1^*\left(\alpha_0\right)$ where $K_1^*\left(\alpha_0\right)$ is the solution of $\Phi\left(K_1 - b\right) + \Phi\left(K_1 - a\right) = \alpha_0$.

We wish to study the power of the test i.e. the value of $\alpha\left(\theta, \mathrm{K}_1, \mathrm{K}_2\right)$ as a function of θ when $\theta \notin \mathrm{H}_0$. Since the situation is symmetric about $\frac{a+b}{2}$, it is sufficient to study the value when $\theta > \mathrm{b}\sigma$. To simplify the formulae, we use the notation $\lambda = \frac{\theta}{\sigma} - \mathrm{b}$ giving $\theta = \lambda\sigma + \mathrm{b}\sigma$ with $\lambda \geq 0$.

For the power of the symmetric test when $\theta > \mathrm{b}\sigma$, we have

$$\alpha\left(\theta, K_1^*, a+b-K_1^*\right) = 2 - \Phi\left(-K_1^* + b + \lambda\right) - \Phi\left(a - K_1^* - \lambda\right)$$

and the probability of Type II error for $\lambda \geq 0$ is given by

$$\beta\left(\lambda, K_1^*, a+b-K_1^*\right) = 1 - \Phi\left(K_1^* - b - \lambda\right) - \Phi\left(K_1^* - a + \lambda\right).$$

Values of $\beta\left(\lambda, K_1^*, a+b-K_1^*\right)$ for different values of $\lambda \geq 0$.

In Tables 2.1 – 2.5 below, the same values of $a=0$, b and α_0^* as in Tables 1.1 – 1.5, are used again. For different values of $\lambda \geq 0$, the probability of Type II error β is computed from,

$$\beta\left(\lambda, K_1^*, a+b-K_1^*\right) \ = \ 1-\Phi\left(K_1^*-b-\lambda\right)-\Phi\left(K_1^*-a+\lambda\right).$$

Table 2.1 $\alpha_0^*=0.1$						
Value of 'b' →	0,0000	0,2500	0,5000	1,0000	2,0000	3,0000
$K_1^* \rightarrow$	-1,6449	-1,5327	-1,4456	-1,3388	-1,2845	-1,2816
$K_2^*=a+b-K_1^*$	1,6449	1,7827	1,9456	2,3388	3,2845	4,2816
β **when** $\lambda=1\rightarrow$	0,7364	0,7002	0,6704	0,6322	0,6120	0,6109
β **when** $\lambda=2\rightarrow$	0,3611	0,3201	0,2896	0,2542	0,2371	0,2363
β **when** $\lambda=3\rightarrow$	0,0877	0,0711	0,0600	0,0483	0,0431	0,0429
β **when** $\lambda=4\rightarrow$	0,0093	0,0068	0,0053	0,0039	0,0033	0,0033
β **when** $\lambda=5\rightarrow$	0,0004	0,0003	0,0002	0,0001	0,0001	0,0001

Table 2.2 $\alpha_0^*=0.05$						
Value of 'b' →	0,0000	0,2500	0,5000	1,0000	2,0000	3,0000
$K_1^* \rightarrow$	-1,9600	-1,8502	-1,7697	-1,6815	-1,6461	-1,6449
$K_2^*=a+b-K_1^*$	1,9600	2,1002	2,2697	2,6815	3,6461	4,6449
β **when** $\lambda=1\rightarrow$	0,8299	0,8014	0,7787	0,7521	0,7409	0,7405
β **when** $\lambda=2\rightarrow$	0,4840	0,4404	0,4089	0,3750	0,3617	0,3612
β **when** $\lambda=3\rightarrow$	0,1492	0,1251	0,1093	0,0937	0,0879	0,0877
β **when** $\lambda=4\rightarrow$	0,0207	0,0158	0,0129	0,0102	0,0093	0,0093
β **when** $\lambda=5\rightarrow$	0,0012	0,0008	0,0006	0,0005	0,0004	0,0004

Table 2.3 $\alpha_0^* = 0.02$						
Value of 'b' →	0,0000	0,2500	0,5000	1,0000	2,0000	3,0000
$K_1^* \rightarrow$	-2,3263	-2,2193	-2,1462	-2,0759	-2,0543	-2,0538
$K_2^* = a + b - K_1^*$	2,3263	2,4693	2,6462	3,0759	4,0543	5,0538
β **when** $\lambda = 1 \rightarrow$	0,9072	0,8884	0,8740	0,8590	0,8541	0,8540
β **when** $\lambda = 2 \rightarrow$	0,6279	0,5868	0,5581	0,5303	0,5216	0,5214
β **when** $\lambda = 3 \rightarrow$	0,2503	0,2175	0,1966	0,1777	0,1721	0,1720
β **when** $\lambda = 4 \rightarrow$	0,0471	0,0375	0,0319	0,0272	0,0258	0,0258
β **when** $\lambda = 5 \rightarrow$	0,0038	0,0027	0,0022	0,0017	0,0016	0,0016

Table 2.4 $\alpha_0^* = 0.01$						
Value of 'b' →	0,0000	0,2500	0,5000	1,0000	2,0000	3,0000
$K_1^* \rightarrow$	-2,5758	-2,4707	-2,4022	-2,3422	-2,3266	-2,3263
$K_2^* = a + b - K_1^*$	2,5758	2,7207	2,9022	3,3422	4,3266	5,3263
β **when** $\lambda = 1 \rightarrow$	0,9423	0,9292	0,9195	0,9102	0,9077	0,9076
β **when** $\lambda = 2 \rightarrow$	0,7176	0,6811	0,6562	0,6339	0,6280	0,6279
β **when** $\lambda = 3 \rightarrow$	0,3357	0,2983	0,2750	0,2553	0,2504	0,2503
β **when** $\lambda = 4 \rightarrow$	0,0772	0,0631	0,0550	0,0487	0,0471	0,0471
β **when** $\lambda = 5 \rightarrow$	0,0077	0,0057	0,0047	0,0039	0,0038	0,0038

Table 2.5 $\alpha_0^* = 0.005$						
Value of 'b' →	0,0000	0,2500	0,5000	1,0000	2,0000	3,0000
$K_1^* \rightarrow$	-2,8070	-2,7036	-2,6393	-2,5876	-2,5760	-2,5758
$K_2^* = a + b - K_1^*$	2,8070	2,9536	3,1393	3,5876	4,5760	5,5758
β **when** $\lambda = 1 \rightarrow$	0,9646	0,9557	0,9494	0,9438	0,9425	0,9425
β **when** $\lambda = 2 \rightarrow$	0,7902	0,7592	0,7387	0,7216	0,7177	0,7176
β **when** $\lambda = 3 \rightarrow$	0,4235	0,3835	0,3592	0,3400	0,3358	0,3357
β **when** $\lambda = 4 \rightarrow$	0,1164	0,0974	0,0868	0,0789	0,0772	0,0772
β **when** $\lambda = 5 \rightarrow$	0,0142	0,0108	0,0091	0,0079	0,0077	0,0077

Reading horizontally any row in these tables, the value of β does not change dramatically as b changes. For example, a mean value θ which is 2σ larger than the null hypothesis limit $b\sigma$ i.e. $\lambda = 2$, has a probability of non-detection that is not heavily affected by the value of b.

Looking down any row of these tables we see that the non-detection probability β is large for θ values near to $b\sigma$ (e.g. $\lambda = 1$) and becomes quite small when the θ value is 5σ above $b\sigma$. This is true for all values of α_0^* though of course having α_0^* smaller increases all error probabilities. Large values of α_0^* are unlikely to be used in practice.

Note that if $\alpha_0^* \leq 0.02$, the non-detection probability for $\lambda = 2$ is quite large i.e. $\beta > 0.5$ in these examples. In other words, if $\theta = \mathrm{b}\sigma + 2\sigma$ there is a probability of non-detection that is not negligible.

5.6 Summary of Duality Results for Minimax Tests

The minimization of $\sup\limits_{-\infty<\theta<+\infty} R\left(\theta, d\right)$ with respect to K_1 and K_2, leads to a simple solution, that is easy to apply and has a number of interesting properties. The result obtained shows that the optimal choice of K_1 and K_2 always lie on the line $K_1 + K_2 = a + b$. In other words, the acceptance region $\left[K_1\sigma, K_2\sigma\right]$ for the minimax decision rule, is always symmetrically placed relative to the composite hypothesis $\left[a\sigma, b\sigma\right]$. This is not surprising given the intrinsic symmetry of the problem.

The K_1 and K_2 values for the minimax decision rule are $\left(K_1^*, a+b - K_1^*\right)$ where $K_1^*\left(\frac{Q}{1+Q}\right)$ is the solution of,

$$\Phi\left(K_1 - a\right) + \Phi\left(K_1 - b\right) = \frac{Q}{1+Q}$$

and $Q = \dfrac{p_2}{p_1}$ is the penalty ratio defined earlier.

This equation for K_1^* is readily solvable and hence the method is easy to apply given specific values for a, b and Q .

The optimal test rule $\left(K_1^*, a+b-K_1^*\right)$ has a number of characteristic properties,

1. Hedging Risk: For each minimax test, the associated values of α_d^* and β_d^* will always be in the same ratio as p_1 and p_2 i.e. any minimax test will satisfy $\alpha_d^* = Q\beta_d^*$. In other words, a minimax decision rule balances the two suprema of error probabilities in proportion to their penalty costs. This has been proved in Sections 2 and 4 of this report.

2. Dual Formulations: The decision theory formulation in terms of loss function and minimax principle is completely equivalent to alternative formulations which at first sight appears quite different. We saw earlier that the rule $\left(K_1^*, a+b-K_1^*\right)$ as well as being a minimax solution, has three other characterising properties. The rule $\left(K_1^*, a+b-K_1^*\right)$ is,

- the unique **symmetric** rule having $\alpha^*\left(K_1, K_2\right) = \alpha_0$ (Section 2.5.1).
- the unique rule with $\alpha^*\left(K_1, K_2\right) = \alpha_0$ and minimum value of $\tilde{K}_2\left(K_1\right) - K_1$ (Section 2.5.3.7).
- the unique rule with $\alpha^*\left(K_1, K_2\right) = \alpha_0$ and minimum value of $\beta^*\left(K_1, K_2\right)$ (Section 3.4).

An alternative equivalent formulation therefore is to focus the maximum false alarm probability α_d^* as a criterion which the decision maker desires to have fixed at some small value α_0 and look for $\mathbf{d}$ being a symmetric rule and having $\alpha_d^* = \alpha_0$. The solution of this different formulation is simply $\left(K_1^*, a+b-K_1^*\right)$ where $K_1^*(\alpha_0)$ is the solution of $\Phi\left(K_1-a\right)+\Phi\left(K_1-b\right)=\alpha_0$ which is of course the equation associated with the minimax formulation.

Hence we have that for all practical purposes, the choice of p_1 and p_2 and the minimax approach is equivalent to choosing a value for α_0 and choosing K_1 and K_2 such that $K_1+K_2=a+b$. The link between the two formulations is provided by $\alpha_0=\frac{Q}{1+Q}$ or equivalently $\alpha_0=\frac{p_2}{p_1+p_2}$.

Note that either of the other two equivalent characterisations (combining $\alpha^*\left(K_1, K_2\right)=\alpha_0$ with the smallest value of $\tilde{K}_2\left(K_1\right)-K_1$ or combining $\alpha^*\left(K_1, K_2\right)=\alpha_0$ with the smallest value of $\beta^*\left(K_1, K_2\right)$), could be taken as defining the alternative formulation.

The minimax formulation, with penalties p_1 and p_2 defining the loss function, shows how these tests can be derived as decision theory. The formulation emphasizing the maximum false alarm probability is more appealing to safeguards inspectors who will have less difficult in choosing a value for α_0 (a maximum false alarm probability they are willing to tolerate) than in imaging values for p_1 and p_2 .

Appendix: A Property of Gaussian Integrals

(these results are used in Section 4.1.5)

In this appendix we consider in the region $K_2 \geq K_1$, the solutions (K_1, K_2) of the equation,

$$\Phi(K_2 - a) - \Phi(K_1 - a) = \Phi(K_2 - b) - \Phi(K_1 - b)\ .$$[25]

We will show that the only possible solutions, when $b > a$, are given by any (K_1, K_2) having $K_2 = K_1$ and by any (K_1, K_2) having $K_2 > K_1$ and $K_1 + K_2 = a + b$.

Solutions having $K_2 > K_1$: We first show that when $K_2 > K_1$ and $b > a$, (K_1, K_2) satisfies,

$$\Phi(K_2 - a) - \Phi(K_1 - a) = \Phi(K_2 - b) - \Phi(K_1 - b)$$

if and only if $K_1 + K_2 = a + b$.

Proof: To see this we let

$$\Psi(x) = \Phi(K_2 - x) - \Phi(K_1 - x) = \frac{1}{\sqrt{2\pi}} \int_{K_1 - x}^{K_2 - x} e^{-\frac{1}{2}u^2} du\ .$$

The function $\Psi(x)$ has the following properties,

(i) $\Psi(x)$ is symmetric about the point $x = \frac{K_1 + K_2}{2}$,

---

[25] Where $\Phi(K) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{K} e^{-\frac{1}{2}u^2} du$

(ii) $\Psi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(K_2 - x)^2} \left\{ e^{\frac{1}{2}(K_2 - K_1)(K_2 + K_1 - 2x)} - 1 \right\}$

(iii) $\Psi(x)$ has a unique stationary point at $x = \frac{K_1 + K_2}{2}$,

(iv) for $\forall x < \frac{K_1 + K_2}{2}$, $\Psi'(x) > 0$,

(v) for $\forall x > \frac{K_1 + K_2}{2}$, $\Psi'(x) < 0$.

Hence we have $\Psi(x) = \Psi(y)$ if and only if $\left| x - \frac{K_1 + K_2}{2} \right| = \left| y - \frac{K_1 + K_2}{2} \right|$

i.e. equality if and only if $x = y$ or $x - \frac{K_1 + K_2}{2} = -\left(y - \frac{K_1 + K_2}{2} \right)$.

Hence if $x \neq y$, $\Psi(x) = \Psi(y)$ implies $x - \frac{K_1 + K_2}{2} = -\left(y - \frac{K_1 + K_2}{2} \right)$.

But $x - \frac{K_1 + K_2}{2} = -\left(y - \frac{K_1 + K_2}{2} \right)$ is the same as $x + y = K_1 + K_2$.

Hence putting $x = a$ and $y = b$ and since $b > a$, we have that

$$\frac{1}{\sqrt{2\pi}} \int_{K_1 - a}^{K_2 - a} e^{-\frac{1}{2}u^2} du = \frac{1}{\sqrt{2\pi}} \int_{K_1 - b}^{K_2 - b} e^{-\frac{1}{2}u^2} du$$

if and only if $a + b = K_1 + K_2$. QED[26]

Note: It is straightforward to show that any (K_1, K_2) satisfying $K_1 + K_2 = a + b$ will ensure equality of the integrals. The equation

[26] With thanks to Tim for the short proof of above by using $\Psi(x)$.

$K_1 + K_2 = a + b$ is equivalent to $K_1 - a = -(K_2 - b)$ and is also equivalent to $K_1 - b = -(K_2 - a)$. Substitution of these two in the equation

$$\Phi(K_2 - a) - \Phi(K_1 - a) = \Phi(K_2 - b) - \Phi(K_1 - b)$$

(and using $\Phi(-u) = 1 - \Phi(u)$), gives immediately that any $(\mathrm{K}_1, \mathrm{K}_2)$ satisfying $K_1 + K_2 = a + b$ will satisfy the equation. It is the converse that is less evident.

Solutions having $K_2 = K_1$: It is obvious that any $(\mathrm{K}_1, \mathrm{K}_2)$ satisfying $K_2 = K_1$ is a solution (the range of integration is zero on both sides of the equation). Conversely having the range of integration equal to zero on both sides, implies $K_2 = K_1$.

Conclusion: Hence we have shown in this appendix that when $b > a$, the set of possible solutions $(\mathrm{K}_1, \mathrm{K}_2)$ of,

$$\frac{1}{\sqrt{2\pi}} \int_{K_1 - a}^{K_2 - a} e^{-\frac{1}{2}u^2} du = \frac{1}{\sqrt{2\pi}} \int_{K_1 - b}^{K_2 - b} e^{-\frac{1}{2}u^2} du$$

is given by two possible types of solution, firstly all $(\mathrm{K}_1, \mathrm{K}_2)$ having $K_2 = K_1$ (trivial solutions) and secondly all $(\mathrm{K}_1, \mathrm{K}_2)$ having $K_2 > K_1$ and $K_1 + K_2 = a + b$. No other solutions are possible.

Appendix B: Symbol Glossary

Notation Introduced in Section 1

σ also σ_{MUF} : denotes the standard deviation of the material balance value.

$\left[a\sigma, b\sigma\right]$ where $a \le b$: denotes the range of acceptable values of θ. This set of acceptable values will also be denoted H_0.

$\mathrm{p}_1 > 0$, $\mathrm{p}_2 > 0$: penalties associated with Type 1 error and Type 2 error respectively (i.e. cost of false alarm and cost of non detection)

$\mathrm{Q} = \frac{\mathrm{p}_2}{\mathrm{p}_1}$;

$\alpha_d(\theta)$: probability of rejecting H_0 when θ is true.

α_d^* : supremum of $\alpha_d(\theta)$ over $\theta \in H_0$. α_d^* is commonly referred to as the **size** of the test **d**.

β_d^* : supremum of $1-\alpha_d(\theta)$ over $\theta \notin H_0$

$R(\theta, d)$: expected loss when θ is true and decision rule **d** is used

$\left[K_1\sigma, K_2\sigma\right]$: acceptance region of decision rule where $K_1 \le K_2$

Notation Introduced in Section 2

$\alpha\left(\theta, K_1, K_2\right)$: same as $\alpha_d\,(\theta)$ where **d** has acceptance region $\left[K_1\sigma, K_2\sigma\right]$.

$\alpha^*\left(K_1, K_2\right)$: same as α_d^* where **d** has acceptance region $\left[K_1\sigma, K_2\sigma\right]$.

$\beta^*\left(K_1, K_2\right)$: same as β_d^* where **d** has acceptance region $\left[K_1\sigma, K_2\sigma\right]$.

$\tilde{K}_2\left(K_1\right)$ also written $\tilde{K}_2\left(K_1, \alpha_0\right)$: implicit function defined by $\alpha^*\left(K_1, K_2\right) = \alpha_0$.

K_1^* also written $K_1^*\left(\alpha_0\right)$: the K_1 coordinate of the point where $\alpha^*\left(K_1, K_2\right) = \alpha_0$ i.e. $\tilde{K}_2\left(K_1, \alpha_0\right)$, intersects $K_1 + K_2 = a + b$.

K_1^{*b} also $K_1^{*b}\left(\alpha_0\right)$: the K_1 coordinate of the point where $\alpha^*\left(K_1, K_2\right) = \alpha_0$ i.e. $\tilde{K}_2\left(K_1, \alpha_0\right)$, intersects $K_1 + K_2 = 2b$.

K_1^0 also $K_1^0\left(\alpha_0\right)$: the K_1 value defining the asymptote $K_1 = K_1^0$ of $\tilde{K}_2\left(K_1, \alpha_0\right)$. K_1^0 is defined by equation $\Phi\left(K_1^0 - a\right) = \alpha_0$.

"Symmetric rule" : a rule whose acceptance region $\left[K_1\sigma, K_2\sigma\right]$ satisfies $K_1 + K_2 = a + b$.

Notation Introduced in Section 3

$\ddot{K}_2(K_1)$ also $\ddot{K}_2(K_1, \beta_0)$: implicit function defined by $\beta^*(K_1, K_2) = \beta_0$.

K_1' also $K_1'(\beta_0)$: the K_1 coordinate of point where $\beta^*(K_1, K_2) = \beta_0$ intersects $K_1 + K_2 = a + b$. It is proved that $K_1'(\beta_0) = K_1^*(1 - \beta_0)$.

$K_1^{'b}$ also written $K_1^{'b}(\beta_0)$ the K_1 coordinate of the point where $\beta^*(K_1, K_2) = \beta_0$ intersects $K_1 + K_2 = 2b$.

It is proved that $K_1^{'b}(\beta_0) = b - \Phi^{-1}\left(\frac{\beta_0 + 1}{2}\right)$

Notation Introduced in Section 4

"gorge" also "Q-gorge" : the set of points (K_1, K_2) satisfying $\left\{(K_1, K_2): \alpha^*(K_1, K_2) = Q\beta^*(K_1, K_2)\right\}$ also $\alpha^* = Q\beta^*$.

$\hat{K}_2(K_1)$ also $\hat{K}_2(K_1, Q)$: the implicit function defined by the Q-gorge $\alpha^* = Q\beta^*$.

K_1^{**} also written $K_1^{**}(\mathrm{Q})$: the K_1 coordinate of the point where the Q-gorge $\alpha^* = \mathrm{Q}\,\beta^*$ intersects $K_1 + K_2 = 2b$ **.**

$K_1^G(K_1', K_2')$ also written $K_1^G(K_1', K_2', \mathrm{Q})$**:** the K_1 coordinate of point where the gorge $\alpha^* = \mathrm{Q}\,\beta^*$ meets the line $K_1 + K_2 = K_1' + K_2'$.
$\left(\mathrm{K}_1', \mathrm{K}_2'\right)$ is an arbitrary point .

Notation Introduced in Section 5

"MUF": material balance referred to as "material unaccounted for" or MUF .

"MUF component": the amounts of material in BI, R, S and EI that are used in computing the balance (see section 5.1).

M_{ik} : represents the **true mass** of the k^{th} item in the i^{th} MUF component of the balance accounts;

Z_{ik} : represents the **accounting mass value** for the k^{th} item in the i^{th} MUF component of the balance accounts ;

$\mathrm{L}_{ik} = \mathrm{Z}_{ik} - \mathrm{M}_{ik}$: is the **accounting discrepancy** for the k^{th} item in the i^{th} MUF component of the balance accounts ;

$\mathrm{MUF}_{\mathrm{TRUE}}$: the **true material balance** computed using true values of the material M_{ik} instead of the accountancy values Z_{ik}.

L_{MUF} : the aggregate of all accounting discrepancies L_{ik} in the balance.

"material balance identity": balance expressed as sum of true balance plus aggregate of accounting discrepancies i.e. $MUF = MUF_{TRUE} + L_{MUF}$

Notation for Incomplete Accounts:

"incomplete balance": a balance in which **the true balance,** using the true values M_{ik} of the sets of material referred to in the accounts, **is not zero**.

MUF^* : The star denotes that the balance is from an incomplete balance.

$\mathrm{MUF}^*_{\mathrm{TRUE}}$: the **true material balance** computed using true values M_{ik} instead of the accountancy values Z_{ik} and here referring to an incomplete material balance.

L^*_{MUF} : the aggregate of all accounting discrepancies L_{ik} in the incomplete material balance MUF^* .

"material balance identity" i.e. $MUF^* = MUF^*_{TRUE} + L^*_{MUF}$: incomplete balance expressed as sum of true balance (of material referred to in the accounts) plus aggregate of accounting discrepancies (of material referred to in the accounts).

σ^*_{MUF} : denotes the standard deviation of the value MUF^* of an incomplete material balance.

$\left[a\,\sigma^*_{MUF}\,,\; b\,\sigma^*_{MUF}\right]$: notation for null hypothesis when testing an incomplete balance.

Notation for Accounts having Measurement Bias

B_{ik} : denotes the measurement error expected value of L_{ik} when this expected value is non-zero.

η_{ik} : denotes the measurement error component of L_{ik} having a probability distribution with expected value zero.

L_{MUF} : in accounts with bias, the aggregate discrepancy can be represented as aggregate effect of bias plus an aggregate component having a probability distribution with expected value zero i.e. $L_{MUF} = B_{MUF} + \eta_{MUF}$.

"material balance identity" : in accounts with bias, the balance can be represented as the sum of a true MUF, an aggregate effect of bias and an aggregate component having a probability distribution with expected value zero i.e. $MUF = MUF_{TRUE} + B_{MUF} + \eta_{MUF}$.

$K_1^*(\alpha_0, b)$: same as $K_1^*(\alpha_0)$ but being considered as a function of b.

European Commission

EUR 23800 EN – Joint Research Centre – Institute for the Protection and Security of the Citizen
Title: **Minimax Tests of Composite Null Hypotheses Applied To Nuclear Material Balances**

Author: M. T. Franklin

Luxembourg: Office for Official Publications of the European Communities
2009 – 173 pp. – 21 x 29.7 cm
EUR – Scientific and Technical Research series – ISSN 1018-5593

Abstract: This report describes the application of minimax tests to the assessment of a nuclear material balance. The approach allows safeguards authorities and nuclear materials accountants to assess a balance taking realistic account of operating contingencies in facilities processing nuclear material. The contingencies handled are: false alarm risk (composite null hypothesis test size), the effect of measurement variability, unmeasured hold-up and uncorrected bias. The report describes the numerical solution for choosing the test acceptance region in terms of: the tolerance limits for hold-up and bias, the material balance measurement error standard deviation and the desired test size. The numerical method is easy to use. The method is applicable to balance closure under international safeguards as well to balance closure as part of the inventory control program in a process area.

Keywords: material balance test, uncorrected bias, hold-up.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.